

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
23 December 2004 (23.12.2004)

PCT

(10) International Publication Number  
**WO 2004/111224 A1**

(51) International Patent Classification <sup>2</sup> :	C12N 9/58	(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EU, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NL, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SE, SL, SY, TI, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
(21) International Application Number:	PCT/DK2004/000836	(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GML, KE, LS, MW, MZ, NA, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TI, TM), European (AT, BE, BG, CH, CX, CZ, DE, DK, EL, ES, FI, FR, GB, GR, HU, IE, PT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BE, BJ, CI, CG, CI, CM, GA, GN, IQ, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date:	21 June 2003 (21.06.2003)	
(25) Filing Language:	English	
(26) Publication Language:	English	
(38) Priority Data:		
PA 2003 00916	19 June 2003 (19.06.2003)	DK
PA 2003 01847	12 December 2003 (12.12.2003)	DK
PA 2004 00334	1 March 2004 (01.03.2004)	DK
(71) Applicant (for all designated States except US):	NOVOZYMES A/S [DK/DK]; Krogshøjvej 36, DK-2880 Bagsværd (DK).	
(72) Inventor; and		
(73) Inventor/Applicant (for US only):	LASSEN, Søren, Flensborg [DK/DK]; Gydebekken 38, DK-3520 Farum (DK).	
(74) Common Representative:	NOVOZYMES A/S; Krogshøjvej 36, DK-2880 Bagsværd (DK).	

## Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidelines Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A1

2004/111224

WO

(54) Title: IMPROVED PROTEASES AND METHODS FOR PRODUCING THEM

(57) Abstract: A secreted mature polypeptide derived from an R2A or S1E protease which after maturation has protease activity, which polypeptide when expressed and before maturation comprises a heterologous pro-region.

**TITLE:** Improved proteases and methods for producing them**FIELD OF INVENTION**

A number of microbially derived related proteases are notably difficult to produce in industrially relevant yields, they may be prone to various types of degradation and/or instabilities. The present invention provides methods for producing such proteases by expressing them as proteases comprising a heterologous pro-sequence. The invention further provides the resulting proteases comprising such heterologous prosequences.

The present invention relates to isolated polypeptides having protease activity related to a *Nocardiopsis* sp. protease, and isolated nucleic acid sequences encoding such proteases. The invention furthermore relates to nucleic acid constructs, vectors, and host cells comprising these nucleic acid sequences as well as methods for producing and using the proteases, in particular within animal feed.

**BACKGROUND**

Polypeptides having protease activity, or proteases, are sometimes also designated peptidases, proteinases, peptide hydrolases, or proteolytic enzymes. Proteases may be of the exo-type that hydrolyses peptides starting at either end thereof, or of the endo-type that act internally in polypeptide chains (endopeptidases). Endopeptidases show activity on N- and C-terminally blocked peptide substrates that are relevant for the specificity of the protease in question.

The term "protease" is defined herein as an enzyme that hydrolyses peptide bonds. It includes any enzyme belonging to the EC 3.4 enzyme group (including each of the thirteen subclasses thereof). The EC number refers to Enzyme Nomenclature 1992 from NC-IUBMB, Academic Press, San Diego, California, including supplements 1-5 published in Eur. J. Biochem. 1994, 223, 1-5; Eur. J. Biochem. 1995, 232, 1-6; Eur. J. Biochem. 1996, 237, 1-5; Eur. J. Biochem. 1997, 260, 1-6; and Eur. J. Biochem. 1999, 264, 610-650; respectively. The nomenclature is regularly supplemented and updated; see e.g. the World Wide Web (WWW) at <http://www.chem.qmw.ac.uk/iubmb/enzyme/index.html>.

US patent publication No. 2002/0182672A1 discloses, that if one or two of the last two amino acids at the C-terminus of a polypeptide is/are charged polar: D or E (negatively charged) or K, R, or H (positively charged), the tail would be considered polar, charged, and this makes the polypeptide resistant against proteolytic degradation by a subclass of proteases that recognize non-polar C-terminal tails of secreted proteins.

Another disclosure reported, that proline residues at the C-terminus of nascent polypeptide chains induce degradation of the polypeptide (2002. Prolin residues at the C terminus of nascent chains induce SsrA tagging during translation termination. J.Biol.Chem. 277:33825-33823).

5

## SUMMARY OF THE INVENTION

It is a well-known problem in the art of expressing polypeptides having proteolytic activity, that many of such polypeptides are inherently unstable, they may be subject to autoproteolysis, or they may be targeted for degradation by other proteases already during their production, resulting in sub-optimal yields. Many other factors may contribute to their instability, not all of which are understood at present. It is of great interest to provide proteolytic polypeptides with an increased stability, that may thus be produced in higher yields.

Secreted proteases of the S2A and/or S1E classification, often have a pro-region which is cleaved off from the protease to produce the mature part of the protease. The present inventors have found, that production of S2A and/or S1E proteases as fusion polypeptides comprising a heterologous pro-region results in much improved yields when compared with production of the unaltered wild-type proteases.

Accordingly, in a first aspect the invention relates to a secreted mature polypeptide which after maturation has protease activity, which polypeptide when expressed and before maturation comprises a heterologous pro-region, and which polypeptide:

- (a) comprises an amino acid sequence which is at least 70% identical, %, or preferably 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide shown in SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45;
- (b) comprises an amino acid sequence which is at least 70% identical, %, or preferably 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide encoded by the polynucleotide in SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 26; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 36; SEQ ID NO: 40; or SEQ ID NO: 44;
- (c) comprises a mature part which is a variant of the mature part of the polypeptide having the amino acid sequence of SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37;

SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45, the segment comprising a substitution, deletion, extension, and/or insertion of one or more amino acids;

(d) is encoded by a nucleic acid sequence which hybridizes under very low, low, medium-low, medium, medium-high, high, or very high stringency conditions with:

- 5 (i) a polynucleotide encoding the mature part of a protease, said polynucleotide obtainable from genomic DNA from *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 by use of primers SEQ ID NO's: 26 and 27; from *Nocardiopsis Alba* DSM 15647 by use of primers SEQ ID NO's: 34 and 35; from *Nocardiopsis prasina* DSM 15548 by use of primers SEQ ID NO's: 38 and 39; or from *Nocardiopsis prasina* DSM 15649 by use of primers SEQ ID NO's: 42 and 39;
- 10 (ii) the polynucleotide of SEQ ID NO: 1; of SEQ ID NO: 2; of SEQ ID NO: 25; of SEQ ID NO: 31; of SEQ ID NO: 32; of SEQ ID NO: 36; of SEQ ID NO: 40; or of SEQ ID NO: 44;
- 15 (iii) a subsequence of (i) or (ii) of at least 500 nucleotides, preferably 400, 300, 200, or 100 nucleotides, or
- (iv) a complementary strand of (i), (ii), or (iii);

(e) is an allelic variant of (a), (b), (c), or (d); or

(f) is a fragment of (a), (b), (c), (d), or (f).

20 In a second aspect, the invention relates to an isolated polynucleotide encoding a polypeptide as defined in the first aspect; or an isolated polynucleotide encoding a fusion polypeptide which after maturation has protease activity, which before maturation comprises a heterologous pro-region, and which polypeptide:

(a) comprises an amino acid sequence which is at least 70% identical, %, or preferably 25 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide shown in SEQ ID NO: 26; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45;

(b) comprises an amino acid sequence which is at least 70% identical, %, or preferably 30 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide encoded by the polynucleotide in SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 25; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 36; SEQ ID NO: 40; or SEQ ID NO: 44;

- (c) comprises a mature part which is a variant of the mature part of the polypeptide having the amino acid sequence of SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45, the segment comprising a substitution, deletion, extension, and/or insertion of one or more amino acids;
- 5 (d) is encoded by a nucleic acid sequence which hybridizes under very low, low, medium-low, medium, medium-high, high, or very high stringency conditions with:
- (I) a polynucleotide encoding the mature part of a protease, said polynucleotide obtainable from genomic DNA from *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 by use of primers SEQ ID NO's: 26 and 27; from *Nocardiopsis Alba* DSM 15647 by use of primers SEQ ID NO's: 34 and 35; from *Nocardiopsis prasina* DSM 15648 by use of primers SEQ ID NO's: 38 and 39; or from *Nocardiopsis prasina* DSM 15649 by use of primers SEQ ID NO's: 42 and 39;
- 10 (II) the polynucleotide of SEQ ID NO: 1; of SEQ ID NO: 2; of SEQ ID NO: 25; of SEQ ID NO: 31; of SEQ ID NO: 32; of SEQ ID NO: 36; of SEQ ID NO: 40; or of SEQ ID NO: 44;
- 15 (III) a subsequence of (I) or (II) of at least 500 nucleotides, preferably 400, 300, 200, or 100 nucleotides, or
- (IV) a complementary strand of (I), (II), or (III);
- 20 (e) is an allelic variant of (a), (b), (c), or (d); or
- (f) is a fragment of (a), (b), (c), (d), or (f).

Still, in a third aspect, the invention relates to a recombinant expression vector or polynucleotide construct comprising a polynucleotide as defined in the previous aspect.

Yet a fourth aspect relates to a recombinant host cell comprising a polynucleotide as defined in the second aspect, or an expression vector or polynucleotide construct as defined in the previous aspect.

In a fifth aspect, the invention also relates to a transgenic plant, or plant part, comprising a polynucleotide as defined in the second aspect, or an expression vector or polynucleotide construct as defined in the third aspect.

The sixth aspect of the invention relates to a transgenic, non-human animal, or products, or elements thereof, comprising a polynucleotide as defined in the second aspect, or an expression vector or polynucleotide construct as defined in the third aspect.

The seventh aspect of the invention relates to a method for producing a polypeptide as defined in the first aspect, the method comprising: (a) cultivating a recombinant host cell as

defined in the fourth aspect, or a transgenic plant or animal as defined in the fifth or sixth aspects, to produce a supernatant comprising the polypeptide, and optionally (b) recovering the polypeptides.

Other aspects of the invention relate to: an animal feed additive comprising at least one polypeptide as defined in the first aspect; and

- (a) at least one fat-soluble vitamin, and/or
- (b) at least one water-soluble vitamin, and/or
- (c) at least one trace mineral;

an animal feed composition having a crude protein content of 50 to 800 g/kg and comprising at least one polypeptide as defined in the first aspect, or at least one feed additive of the previous aspect;

a composition comprising at least one polypeptide as defined in the first aspect, together with at least one other enzyme selected from amongst phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.89); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-.-), phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.5); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6);

a method for using at least one polypeptide as defined in the first aspect, for improving the nutritional value of an animal feed, for increasing digestible and/or soluble protein in animal diets, for increasing the degree of hydrolysis of proteins in animal diets, and/or for the treatment of vegetable proteins, the method comprising including the polypeptide(s) in animal feed, and/or in a composition for use in animal feed;

a method for using at least one polypeptide as defined in the first aspect, comprising including the polypeptide(s) in a detergent formulation.

25

## DETAILED DESCRIPTION OF THE INVENTION

Proteases are classified on the basis of their catalytic mechanism into the following groups: Serine proteases (S), Cysteine proteases (C), Aspartic proteases (A), Metalloproteases (M), and Unknown, or as yet unclassified, proteases (U), see Handbook of Proteolytic Enzymes, A.J.Barrett, N.D.Rawlings, J.F.Woessner (eds), Academic Press (1998), in particular the general introduction part.

Serine proteases are ubiquitous, being found in viruses, bacteria and eukaryotes; they include exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activity. Over 20 families (denoted S1 - S27) of serine proteases have been identified, these being grouped into 6

clans denoted SA, SB, SC, SE, SF, and SG, on the basis of structural similarity and functional evidence (Barrett et al. 1998, *Handbook of proteolytic enzymes*). Structures are known for at least four of the clans (SA, SB, SC and SE), these appear to be totally unrelated, suggesting at least four evolutionary origins of serine peptidases. Alpha-lytic endopeptidases belong to the chymotrypsin (SA) clan, within which they have been assigned to subfamily A of the S2 family (S2A).

Another classification system of proteolytic enzymes is based on sequence information, and is therefore used more often in the art of molecular biology; it is described in Rawlings, N.D. et al., 2002, MEROPS: The protease database. *Nucleic Acids Res.* 30:343-346. The MEROPS database is freely available electronically at <http://www.merops.ac.uk>. According to the MEROPS system, the proteolytic enzymes classified as S2A in 'The Handbook of Proteolytic Enzymes', are in MEROPS classified as 'S1E' proteases (Rawlings ND, Barrett AJ. (1993) Evolutionary families of peptidases, *Biochem. J.* 290:205-218).

In particular embodiments, the proteases of the invention and for use according to the invention are selected from the group consisting of:

- (a) proteases belonging to the EC 3.4.-.- enzyme group;
- (b) Serine proteases belonging to the S group of the above Handbook;
- (c1) Serine proteases of peptidase family S2A;
- (c2) Serine proteases of peptidase family S1E as described in *Biochem.J.* 290:205-218 (1993) and in MEROPS a protease database, release 6.20, March 24, 2003, ([www.merops.ac.uk](http://www.merops.ac.uk)). The database is described in Rawlings, N.D., O'Brien, E. A. & Barrett, A.J. (2002) *MEROPS: the protease database*. *Nucleic Acids Res.* 30, 343-346.

For determining whether a given protease is a Serine protease, and a family S2A protease, reference is made to the above Handbook and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

Protease activity can be measured using any assay, in which a substrate is employed, that includes peptide bonds relevant for the specificity of the protease in question. Assay-pH and assay-temperature are likewise to be adapted to the protease in question. Examples of assay-pH-values are pH 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12. Examples of assay-temperatures are 30, 35, 37, 40, 45, 50, 55, 60, 65, 70, 80, 90, or 95°C.

Examples of protease substrates are casein, such as Azurine-Crosslinked Casein (AZCL-casein). Two protease assays are described in Example 2 herein, either of which can be

used to determine protease activity. For the purposes of this invention, the so-called pNA Assay is a preferred assay.

There are no limitations on the origin of the protease of the invention and/or for use according to the invention. Thus, the term protease includes not only natural or wild-type proteases obtained from microorganisms of any genus, but also any mutants, variants, fragments etc. thereof exhibiting protease activity, as well as synthetic proteases, such as shuffled proteases, and consensus proteases. Such genetically engineered proteases can be prepared as is generally known in the art, e.g. by Site-directed Mutagenesis, by PCR (using a PCR fragment containing the desired mutation as one of the primers in the PCR reactions), or by Random Mutagenesis. The preparation of consensus proteins is described in e.g. EP 897985. The term "obtained from" as used herein in connection with a given source shall mean that the polypeptide encoded by the nucleic acid sequence is produced by the source or by a cell in which the nucleic acid sequence from the source is present. In a preferred embodiment, the polypeptide is secreted extracellularly.

In a specific embodiment, the protease is a low-allergenic variant, designed to invoke a reduced immunological response when exposed to animals, including man. The term immunological response is to be understood as any reaction by the immune system of an animal exposed to the protease. One type of immunological response is an allergic response leading to increased levels of IgE in the exposed animal. Low-allergenic variants may be prepared using techniques known in the art. For example the protease may be conjugated with polymer moieties shielding portions or epitopes of the protease involved in an immunological response. Conjugation with polymers may involve *in vitro* chemical coupling of polymer to the protease, e.g. as described in WO 96/17929, WO 98/30682, WO 98/35026, and/or WO 99/00489. Conjugation may in addition or alternatively thereto involve *in vivo* coupling of polymers to the protease. Such conjugation may be achieved by genetic engineering of the nucleotide sequence encoding the protease, inserting consensus sequences encoding additional glycosylation sites in the protease and expressing the protease in a host capable of glycosylating the protease, see e.g. WO 00/26354. Another way of providing low-allergenic variants is genetic engineering of the nucleotide sequence encoding the protease so as to cause the protease to self-oligomerize, effecting that protease monomers may shield the epitopes of other protease monomers and thereby lowering the antigenicity of the oligomers. Such products and their preparation is described e.g. in WO 96/16177. Epitopes involved in an immunological response may be identified by various methods such as the phage display method described in WO 00/26230 and WO 01/83559, or the random approach described in EP 561907. Once an epitope has been

identified, its amino acid sequence may be altered to produce altered immunological properties of the protease by known gene manipulation techniques such as site directed mutagenesis (see e.g. WO 00/26230, WO 00/26354 and/or WO 00/22103) and/or conjugation of a polymer may be done in sufficient proximity to the epitope for the polymer to shield the epitope.

5 The first aspect of the invention relates to a secreted mature polypeptide which after maturation has protease activity, which polypeptide when expressed and before maturation comprises a heterologous pro-region, and which polypeptide:

- (a) comprises an amino acid sequence which is at least 70% identical, %, or preferably 10 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide shown in SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45;
- (b) comprises an amino acid sequences which is at least 70% identical, %, or preferably 15 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the mature part of the polypeptide encoded by the polynucleotide in SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 25; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 36; SEQ ID NO: 40; or SEQ ID NO: 44;
- (c) comprises a mature part which is a variant of the mature part of the polypeptide 20 having the amino acid sequence of SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45, the segment comprising a substitution, deletion, extension, and/or insertion of one or more amino acids;
- (d) is encoded by a nucleic acid sequence which hybridizes under very low, low, medium-low, medium, medium-high, high, or very high stringency conditions with:
  - 25 (i) a polynucleotide encoding the mature part of a protease, said polynucleotide obtainable from genomic DNA from *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 by use of primers SEQ ID NO's: 26 and 27; from *Nocardiopsis Alba* DSM 15647 by use of primers SEQ ID NO's: 34 and 35; from *Nocardiopsis prasina* DSM 15648 by use of primers SEQ ID NO's: 38 and 39; or from *Nocardiopsis prasina* DSM 30 15649 by use of primers SEQ ID NO's: 42 and 39;
  - (ii) the polynucleotide of SEQ ID NO: 1; of SEQ ID NO: 2; of SEQ ID NO: 26; of SEQ ID NO: 31; of SEQ ID NO: 32; of SEQ ID NO: 36; of SEQ ID NO: 40; or of SEQ ID NO: 44;

(III) a subsequence of (I) or (II) of at least 500 nucleotides, preferably 400, 300, 200, or 100 nucleotides, or

(IV) a complementary strand of (I), (II), or (III);

(c) is an allelic variant of (a), (b), (c), or (d); or

(f) is a fragment of (a), (b), (c), (d), or (f).

For the purposes of the present invention, the degree of identity between two amino acid sequences, as well as the degree of identity between two nucleotide sequences, is determined by the program "align" which is a Needleman-Wunsch alignment (i.e. a global alignment). The program is used for alignment of polypeptide, as well as nucleotide sequences. The default scoring matrix BLOSUM50 is used for polypeptide alignments, and the default identity matrix is used for nucleotide alignments. The penalty for the first residue of a gap is -12 for polypeptides and -16 for nucleotides. The penalties for further residues of a gap are -2 for polypeptides, and -4 for nucleotides.

"Align" is part of the FASTA package version v20uS (see W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448, and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA," Methods in Enzymology 183:63-98). FASTA protein alignments use the Smith-Waterman algorithm with no limitation on gap size (see "Smith-Waterman algorithm", T. F. Smith and M. S. Waterman (1981) J. Mol. Biol. 147:195-197).

The degree of identity between two amino acid sequences may also be determined by the Clustal method (Higgins, 1989, CABIOS 5: 151-153) using the LASERGENE™ MEGALIGN™ software (DNASTAR, Inc., Madison, WI) with an identity table and the following multiple alignment parameters: Gap penalty of 10, and gap length penalty of 10. Pairwise alignment parameters are Ktuple=1, gap penalty=3, windows=5, and diagonals=5. The degree of identity between two nucleotide sequences may be determined using the same algorithm and software package as described above with the following settings: Gap penalty of 10, and gap length penalty of 10. Pairwise alignment parameters are Ktuple=3, gap penalty=3 and windows=20.

A fragment of one of the encoding polynucleotide sequences of the invention is a polynucleotide which encodes a polypeptide having one or more amino acids deleted from the amino and/or carboxyl terminus compared to the full-length amino acid sequence. In one embodiment a fragment encodes at least 75 amino acid residues, or at least 100 amino acid residues, or at least 125 amino acid residues, or at least 150 amino acid residues, or at least

160 amino acid residues, or at least 165 amino acid residues, or at least 170 amino acid residues, or at least 175 amino acid residues.

An allelic variant denotes any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

The present invention also relates to isolated polypeptides having protease activity and which are encoded by nucleic acid sequences which hybridize under very low, or low, or low-medium, medium, medium-high, high, or very high stringency conditions with a nucleic acid probe which hybridizes under the same conditions with (I) a polynucleotide encoding the mature part of a protease obtainable from genomic DNA from *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 by use of primers SEQ ID NO's: 26 and 27; from *Nocardiopsis Alba* DSM 15647 by use of primers SEQ ID NO's: 34 and 35; from *Nocardiopsis prasina* DSM 15648 by use of primers SEQ ID NO's: 38 and 39; or from *Nocardiopsis prasina* DSM 15649 by use of primers SEQ ID NO's: 42 and 39; (II) the polynucleotide of SEQ ID NO: 1; of SEQ ID NO: 2; of SEQ ID NO: 25; of SEQ ID NO: 31; of SEQ ID NO: 32; of SEQ ID NO: 36; of SEQ ID NO: 40; or of SEQ ID NO: 44; (III) a subsequence of (I) or (II) of at least 500 nucleotides, preferably 400, 300, 200, or 100 nucleotides, or (IV) a complementary strand of (I), (II), or (III) (J. Sambrook, E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2nd edition, Cold Spring Harbor, New York). In one particular embodiment the nucleic acid probe is selected from amongst the nucleic acid sequences of (a), (b), or (c) above. A polynucleotide corresponding to the mature peptide encoding part of SEQ ID NO: 1, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 36, SEQ ID NO: 40, or of SEQ ID NO: 44, is a preferred probe.

The nucleic acid sequences of SEQ ID NO: 1, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 36, SEQ ID NO: 40, or SEQ ID NO: 44, or a subsequence thereof, as well as the amino acid sequences of SEQ ID NO: 28; SEQ ID NO: 30; SEQ ID NO: 33; SEQ ID NO: 37, SEQ ID NO: 41, SEQ ID NO: 43, or SEQ ID NO: 45, or a fragment thereof, and even a genomic polynucleotide encoding a protease obtainable from genomic DNA from *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 by use of primers SEQ ID NO's: 26 and 27; from *Nocardiopsis Alba* DSM 15647 by use of primers SEQ ID NO's: 34 and 35; from *Nocardiopsis prasina* DSM 15648 by use of primers SEQ ID NO's: 38 and 39; or from *Nocardiopsis prasina* DSM 15649 by use of primers SEQ ID NO's: 42 and 39, or a subsequence thereof, may be used to design a nucleic acid probe to identify and clone DNA encoding polypeptides having protease

activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic or cDNA of the genus or species of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, preferably at least 25, and more preferably at least 35 nucleotides in length. Longer probes can also be used. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{35}\text{S}$ , biotin, or avidin). Such probes are encompassed by the present invention.

Thus, a genomic DNA or cDNA library prepared from such other organisms may be screened for DNA that hybridizes with the probes described above and which encodes a polypeptide having protease activity. Genomic or other DNA from such other organisms may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA which is homologous with SEQ ID NO: 1 or a subsequence thereof, the carrier material is used in a Southern blot. For purposes of the present invention, hybridization indicates that the nucleic acid sequence hybridizes to a labeled nucleic acid probe corresponding to the nucleic acid sequence shown in SEQ ID NO: 1, its complementary strand, or a subsequence thereof, under very low to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions are detected using X-ray film.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 0.2 x SSC, 0.2% SDS, 20% formamide preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

For short probes about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at 5°C to 10°C below the calculated  $T_m$  using the calculation according to Bolton and McCarthy (1982,

Proceedings of the National Academy of Sciences USA 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures.

5 For short probes about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SSC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated  $T_m$ .

The present invention also relates to variants of the polypeptide of the invention, comprising a substitution, deletion, and/or insertion of one or more amino acids.

10 In a particular embodiment, amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or 15 another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter the specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In, The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Vai, Ala/Glu, and Asp/Gly as well as these in 25 reverse.

In a particular embodiment, the polypeptides of the invention and for use according to the invention are acid-stable. For the present purposes, the term acid-stable means that the residual activity after 2 hours of incubation at pH 3.0 and 37°C, is at least 50%, as compared to the residual activity of a corresponding sample incubated for 2 hours at pH 9.0 and 5°C. In a 30 particular embodiment, the residual activity is at least 60%, 70%, 80% or at least 90%.

In particular embodiments, the polypeptide of the invention is i) a bacterial protease; ii) a protease of the phylum *Actinobacteria*; iii) of the class *Actinobacteria*; iv) of the order *Actinomycetales* v) of the family *Nocardiopsaceae*; vi) of the genus *Nocardiopsis*; and/or a protease derived from vii) *Nocardiopsis* species such as *Nocardiopsis alba*, *Nocardiopsis*

*antarctica*, *Nocardiopsis prasina*, *composta*, *exhalans*, *halophila*, *halotolerans*, *kunshanensis*, *listeri*, *lucentensis*, *metallicus*, *synnemataformans*, *trehalosi*, *tropica*, *umidischolae*, *xinjiangensis*, or *Nocardiopsis dassonvillei*, for example *Nocardiopsis dassonvillei* DSM 43235.

The above taxonomy is according to the chapter: The road map to the Manual by G.M. Garrity & J. G. Holt in Bergey's Manual of Systematic Bacteriology, 2001, second edition, volume 1, David R. Boone, Richard W. Castenholz.

It will be understood that for the aforementioned species, the invention encompasses both the perfect and imperfect states, and other taxonomic equivalents, e.g., anamorphs, regardless of the species name by which they are known. Those skilled in the art will readily recognize the identity of appropriate equivalents.

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL). E.g., *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43235 is publicly available from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). The strain was also deposited at other depositary institutions as follows: ATCC 23219, IMRU 1250, NCTC 10489.

Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) using the above-mentioned probes. Techniques for isolating microorganisms from natural habitats are well known in the art. The nucleic acid sequence may then be derived by similarly screening a genomic or cDNA library of another microorganism. Once a nucleic acid sequence encoding a polypeptide has been detected with the probe(s), the sequence may be isolated or cloned by utilizing techniques which are known to those of ordinary skill in the art (see, e.g., Sambrook et al., 1989, *supra*).

As defined herein, an "isolated" polypeptide is a polypeptide which is essentially free of other polypeptides, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by SDS-PAGE.

Polypeptides encoded by nucleic acid sequences of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleic acid sequence (or a portion thereof) encoding another polypeptide

to a nucleic acid sequence (or a portion thereof) of the present invention. Techniques for producing fusion polypeptides are known in the art, e.g. PCR, or ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator.

In the present context, non-polar amino acids are G, A, V, L, I, M, P, F or W; and uncharged polar amino acids are S, T, N, Q, Y, or C. The terms "non-polar" and "uncharged polar" when used to describe amino acids in a polypeptide are generally recognized in the art as characterizing the side-chain of the amino acid. Hence, for instance, the free carboxylic acid of the C-terminal amino acid in a polypeptide is not considered when determining whether this amino acid is a non-polar or uncharged polar amino acid.

A preferred embodiment relates to a polypeptide of the first aspect which mature part is a wildtype polypeptide; an artificial variant of a wildtype polypeptide said variant having one or more amino-acid(s) added to the C-terminus as compared to the wildtype and preferably the one or more added amino acid(s) is (are) non-polar or uncharged and even more preferably the one or more added amino acid(s) is one or more of Q, S, V, A, or P; a shuffled polypeptide; or a protein-engineered polypeptide.

Yet another preferred embodiment relates to a polypeptide of the first aspect, which comprises at least three non-polar or uncharged polar amino acids within the last four amino acids of the C-terminus of the polypeptide;

Another preferred embodiment relates to a polypeptide of the first aspect, wherein the one or more added amino acids are selected from the group consisting of: QSHV/QSAP, QSAP, QP, TL, TT, QL, TP, LP, TI, IQ, QP, PI, LT, TQ, IT, QQ, and PQ.

The inventors determined, that the polypeptides of the present invention were produced in even greater yields when they were expressed as mature proteases fused to a heterologous pro-region, and/or fused to a heterologous secretion signal peptide, as shown in the examples below.

Accordingly, a preferred embodiment relates to the polypeptide according to the first aspect which when expressed and before maturation comprises a heterologous secretion signal-peptide which is cleaved from the polypeptide when the polypeptide is secreted, preferably the heterologous secretion signal peptide is derived from a heterologous protease, preferably the heterologous secretion signal peptide comprises an amino acid sequence having a sequence identity of at least 70%, or preferably 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, with the amino acid sequence encoded by polynucleotides 1 – 81 of SEQ ID NO: 2, or SEQ ID NO: 44.

Accordingly, a preferred embodiment relates to the polypeptide according to the first aspect which when expressed and before maturation comprises a heterologous pro-region from a protease; preferably the pro-region is derived from an S2A or S1E protease, more preferably the pro-region is encoded by a polynucleotide which hybridizes under very low, low, medium-low, medium, medium-high, high, or very high stringency conditions with a polynucleotide encoding the pro-region shown in position -166 to -1 of SEQ ID NO: 28, in position 1-166 of SEQ ID NO: 30, in position -167 to -1 of SEQ ID NO: 33, in position -166 to -1 of SEQ ID NO: 37, in position -165 to -1 of SEQ ID NO: 41, in position -165 to -1 of SEQ ID NO: 43, in position -165 to -1 of SEQ ID NO: 45, in position 1-165 of SEQ ID NO: 46, in position 1-166 of SEQ ID NO: 47, in position 1-166 of SEQ ID NO: 48, in position 1-166 of SEQ ID NO: 49, in position 1-166 of SEQ ID NO: 50, in position 1-165 of SEQ ID NO: 51, in position 1-166 of SEQ ID NO: 52, or in position 1-166 of SEQ ID NO: 53; and most preferably it is at least 70% identical, or preferably 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, identical to the pro-region shown in position -166 to -1 of SEQ ID NO: 28, in position 1-166 of SEQ ID NO: 30, in position -167 to -1 of SEQ ID NO: 33, in position -165 to -1 of SEQ ID NO: 37, in position -165 to -1 of SEQ ID NO: 41, in position -165 to -1 of SEQ ID NO: 43, in position -165 to -1 of SEQ ID NO: 45, in position 1-165 of SEQ ID NO: 46, in position 1-166 of SEQ ID NO: 47, in position 1-166 of SEQ ID NO: 48, in position 1-166 of SEQ ID NO: 49, in position 1-166 of SEQ ID NO: 50, in position 1-165 of SEQ ID NO: 51, in position 1-166 of SEQ ID NO: 52, or in position 1-166 of SEQ ID NO: 53.

#### Nucleic Acid Sequences

The present invention also relates to isolated nucleic acid sequences that encode a polypeptide of the present invention. Particular nucleic acid sequences of the invention are the polynucleotides of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 36, SEQ ID NO: 40, or SEQ ID NO: 44. Another particular nucleic acid sequence of the invention is the sequence, preferably the mature polypeptide encoding region thereof, which is obtainable from genomic DNA from *Nocardiopsis dassonvillei* subspecies *dassonvillei* DSM 43235. The present invention also encompasses nucleic acid sequences which encode a polypeptide having the amino acid sequence of amino acids shown in positions 1 to 188, or positions -165 to 188, of SEQ ID NO: 43, which nucleic acid sequences differ from the corresponding parts of SEQ ID NO: 1 by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of the above polynucleotides which encode polypeptide fragments that have protease activity.

A subsequence of a polynucleotide is a nucleic acid sequence from which one or more nucleotides from the 5' and/or 3' end has been deleted. Preferably, a subsequence contains at least 225 nucleotides, more preferably at least 300 nucleotides, even more preferably at least 375, 450, 500, 531, 600, 700, 800, 900 or 1000 nucleotides. The present invention also relates to nucleotide sequences which have a degree of identity to the polynucleotide of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 36, SEQ ID NO: 40, or SEQ ID NO: 44, of at least 85%, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or at least 99%.

The techniques used to isolate or clone a nucleic acid sequence encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the nucleic acid sequences of the present invention from such genomic DNA can be effected, e.g., by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) may be used. The nucleic acid sequence may be cloned from a strain of *Nocardiopsis* or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleic acid sequence.

The term "isolated nucleic acid sequence" as used herein refers to a nucleic acid sequence which is essentially free of other nucleic acid sequences, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably at least about 60% pure, even more preferably at least about 80% pure, and most preferably at least about 90% pure as determined by agarose electrophoresis. For example, an isolated nucleic acid sequence can be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the nucleic acid sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host cell where multiple copies or clones of the nucleic acid sequence will be replicated. The nucleic acid sequence may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

Modification of a nucleic acid sequence encoding a polypeptide of the present invention may be necessary for the synthesis of polypeptides substantially similar to the polypeptide. The term "substantially similar" to the polypeptide refers to non-naturally occurring forms of the

polypeptide. These polypeptides may differ in some engineered way from the polypeptide isolated from its native source, e.g., variants that differ in specific activity, thermostability, pH optimum, allergenicity, or the like. The variant sequence may be constructed on the basis of the nucleic acid sequence presented as the polypeptide encoding part of the polynucleotides of the invention, e.g. a subsequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the polypeptide encoded by the nucleic acid sequence, but which correspond to the codon usage of the host organism intended for production of the protease, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence. For a general description of nucleotide substitution, see, e.g., 5 Ford *et al.*, 1991, *Protein Expression and Purification* 2: 95-107. Low-allergenic polypeptides can 10 e.g. be prepared as described above.

It will be apparent to those skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acid residues essential to the activity of the polypeptide encoded by the isolated nucleic acid 15 sequence of the invention, and therefore preferably not subject to substitution, may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (see, e.g., Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, mutations are introduced at every positively charged residue in the molecule, and the resultant mutant molecules are tested for protease activity to identify amino acid residues that 20 are critical to the activity of the molecule. Sites of substrate-protease interaction can also be determined by analysis of the three-dimensional structure as determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labelling (see, e.g., de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *Journal of Molecular Biology* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Letters* 309: 59-64).

The present invention also relates to isolated nucleic acid sequences encoding a polypeptide of the present invention, which hybridize under very low stringency conditions, preferably low stringency conditions, more preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with a nucleic acid probe which hybridizes under the same conditions with the nucleic acid sequence of the invention or its complementary strand; or allelic variants and subsequences thereof (Sambrook *et al.*, 1989, *supra*), as defined herein.

The present invention also relates to isolated nucleic acid sequences produced by (a) hybridizing a DNA under very low, low, medium, medium-high, high, or very high stringency conditions with (i) a polynucleotide of the invention, (ii) a subsequence of (i), or (iii) a complementary strand of (i), or (ii); and (b) isolating the nucleic acid sequence. The subsequence is preferably a sequence of at least 100 nucleotides such as a sequence that encodes a polypeptide fragment which has protease activity.

The introduction of a mutation into the nucleic acid sequence to exchange one nucleotide for another nucleotides may be accomplished by site-directed mutagenesis using any of the methods known in the art. Particularly useful is the procedure that utilizes a supercoiled, double stranded DNA vector with an insert of interest and two synthetic primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, extend during temperature cycling by means of *Pfu* DNA polymerase. On incorporation of the primers, a mutated plasmid containing staggered nicks is generated. Following temperature cycling, the product is treated with *DpnI* which is specific for methylated and hemimethylated DNA to digest the parental DNA template and to select for mutation-containing synthesized DNA. Other procedures known in the art may also be used. The invention also relates to an isolated polynucleotide encoding a polypeptide as defined in the first aspect.

#### Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising a nucleic acid sequence of the present invention operably linked to one or more control sequences that direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences. Expression will be understood to include any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

"Nucleic acid construct" is defined herein as a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid combined and juxtaposed in a manner that would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term expression cassette 5 when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" is defined herein as a nucleic acid sequence that directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by a ribosome binding site 10 (prokaryotes) or by the ATG start codon (eukaryotes) located just upstream of the open reading frame at the 5' end of the mRNA and a transcription terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

An isolated nucleic acid sequence encoding a polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of 15 the nucleic acid sequence prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

The term "control sequences" is defined herein to include all components that are necessary or advantageous for the expression of a polypeptide of the present invention. Each 20 control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcriptional terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of 25 introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide. The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the expression of a polypeptide.

30 The control sequence may be an appropriate promoter sequence, a nucleic acid sequence that is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be

obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (*dagA*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus subtilis* *xylA* and *xylB* genes, and prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), as well as the lac promoter (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242: 74-94; and in Sambrook et al., 1989, supra.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos et al., 1992, Yeast 8: 423-488.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans*

anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

Preferred terminators for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romano et al., 1992, supra.

Preferred terminators for bacterial host cells, such as a *Bacillus* host cell, are the terminators from *Bacillus licheniformis* alpha-amylase gene (amyL), the *Bacillus stearothermophilus* maltogenic amylase gene (amyM), or the *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ).

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention.

Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase.

Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Molecular Cellular Biology* 15: 5983-5990.

The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic

acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to the coding sequence. The foreign signal peptide coding region 5 may be required where the coding sequence does not naturally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearnothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* alpha-amylase, *Bacillus stearnothermophilus* neutral proteases (*nprT*, *nprS*, *nprM*), and *Bacillus subtilis* *prsA*. Further signal peptides are described by Simonen and Palva, 1993, Microbiological 10 Reviews 57: 109-137.

Effective signal peptide coding regions for filamentous fungal host cells are the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos et al., 1992, *supra*.

The control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region 25 may be obtained from the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliothermophila thermophila* laccase (WO 95/33838).

Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems include the *lac*, *lac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the polypeptide would be operably linked with the regulatory sequence.

#### Expression Vectors

The present invention also relates to recombinant expression vectors comprising a nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (e.g., a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

The vector may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated

together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*. Suitable markers for yeast host cells are *ADE2*, *HIS3*, *LEU2*, *LYS2*, *MET3*, *TRP1*, and *URA3*. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hygB* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (crotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), *trpC* (anthranilate synthase), as well as equivalents thereof. Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* gene of *Streptomyces hygroscopicus*.

The vectors of the present invention preferably contain an element(s) that permits stable integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome.

For integration into the host cell genome, the vector may rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAM131 permitting replication in *Bacillus*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes it functioning temperature-sensitive in the host cell (see, e.g., Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1433).

More than one copy of a nucleic acid sequence of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, e.g., Sambrook et al., 1989, *supra*).

The protease may also be co-expressed together with at least one other enzyme of interest for animal feed, such as phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.69); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-~), phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.6); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6).

The enzymes may be co-expressed from different vectors, from one vector, or using a mixture of both techniques. When using different vectors, the vectors may have different selectable markers, and different origins of replication. When using only one vector, the genes can be expressed from one or more promoters. If cloned under the regulation of one promoter (di- or multi-cistronic), the order in which the genes are cloned may affect the expression levels of the proteins. The protease may also be expressed as a fusion protein, i.e. that the gene encoding the protease has been fused in frame to the gene encoding another protein. This protein may be another enzyme or a functional domain from another enzyme.

Accordingly, the invention also relates to a recombinant expression vector or polynucleotide construct comprising a polynucleotide of the invention.

#### Host Cells

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. A vector comprising a nucleic acid sequence of the present invention is introduced into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source. The host cell may be a unicellular microorganism, e.g., a prokaryote, or a non-unicellular microorganism, e.g., a eukaryote.

Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, or a *Streptomyces* cell, or cells of lactic acid bacteria; or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. Lactic acid bacteria include, but are not limited to, species of the genera *Lactococcus*, *Lactobacillus*, *Leuconostoc*, *Streptococcus*, *Pediococcus*, and *Enterococcus*. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, e.g., *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus laetus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis*; or a *Streptomyces* cell, e.g., *Streptomyces lividans* or *Streptomyces mutinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. In a preferred embodiment, the bacterial host cell is a *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* or *Bacillus subtilis* cell. In another preferred embodiment, the *Bacillus* cell is an alkalophilic *Bacillus*.

The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, e.g., Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, e.g., Young and Spizizen, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, e.g., Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, e.g., Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278). The host cell may be a eukaryote, such as a non-human animal cell, an insect cell, a plant cell, or a fungal cell. In one particular embodiment, the host cell is a fungal cell. "Fungi" as used herein

includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al.*, *In, Ainsworth and Bisby's Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*).

5 In another particular embodiment, the fungal host cell is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology 10 and Activities of Yeast* (Skinner, F.A., Pasmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

The yeast host cell may be a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell.

15 The fungal host cell may be a filamentous fungal cell. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, *supra*). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon 20 catabolism may be fermentative.

Examples of filamentous fungal host cells are cells of species of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Mycelopithora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, or *Trichoderma*.

25 Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton *et al.*, 1984, *Proceedings of the National Academy of Sciences USA* 81: 1470-1474. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156 and WO 96/00787. Yeast may be transformed using the procedures 30 described by Becker and Guarante, *In* Abelson, J.N. and Simon, M.J., editors, *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; and Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

The invention relates to a recombinant host cell comprising a polynucleotide of the invention, or an expression vector or polynucleotide construct of the invention. In a preferred embodiment, the recombinant host cell is a *Bacillus* cell.

## 6 Plants

The present invention also relates to a transgenic plant, plant part, or plant cell which has been transformed with a nucleic acid sequence encoding a polypeptide having protease activity of the present invention so as to express and produce the polypeptide in recoverable quantities. The polypeptide may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide may be used as such for improving the quality of a food or feed, e.g., improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive factor.

In a particular embodiment, the polypeptide is targeted to the endosperm storage vacuoles in seeds. This can be obtained by synthesizing it as a precursor with a suitable signal peptide, see Horvath et al in PNAS, Feb. 15, 2000, vol. 97, no. 4, p. 1914-1919.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot) or engineered variants thereof. Examples of monocot plants are grasses, such as meadow grass (blue grass, *Poa*), forage grass such as *festuca*, *lolium*, temperate grass, such as *Agrostis*, and cereals, e.g., wheat, oats, rye, barley, rice, sorghum, and maize (corn). Examples of dicot plants are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family Brassicaceae), such as cauliflower, rape seed, and the closely related model organism *Arabidopsis thaliana*. Low-phytate plants as described e.g. in US patent no. 5,689,064 and US patent no. 6,111,168 are examples of engineered plants.

Examples of plant parts are stem, callus, leaves, root, fruits, seeds, and tubers. Also specific plant tissues, such as chloroplast, apoplast, mitochondria, vacuole, peroxisomes, and cytoplasm are considered to be a plant part. Furthermore, any plant cell, whatever the tissue origin, is considered to be a plant part.

Also included within the scope of the present invention are the progeny of such plants, plant parts and plant cells.

The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with methods known in the art. Briefly, the plant or plant cell is constructed by incorporating one or more expression constructs encoding a polypeptide of the present invention into the plant host genome and propagating the resulting modified plant or plant cell into a transgenic plant or plant cell.

Conveniently, the expression construct is a nucleic acid construct which comprises a nucleic acid sequence encoding a polypeptide of the present invention operably linked with appropriate regulatory sequences required for expression of the nucleic acid sequence in the plant or plant part of choice. Furthermore, the expression construct may comprise a selectable marker useful for identifying host cells into which the expression construct has been integrated and DNA sequences necessary for introduction of the construct into the plant in question (the latter depends on the DNA introduction method to be used).

The choice of regulatory sequences, such as promoter and terminator sequences and optionally signal or transit sequences are determined, for example, on the basis of when, where, and how the polypeptide is desired to be expressed. For instance, the expression of the gene encoding a polypeptide of the present invention may be constitutive or inducible, or may be developmental, stage or tissue specific, and the gene product may be targeted to a specific tissue or plant part such as seeds or leaves. Regulatory sequences are, for example, described by Tague *et al.*, 1988, *Plant Physiology* 86: 506.

For constitutive expression, the 35S-CaMV promoter may be used (Franck *et al.*, 1980, *Cell* 21: 285-294). Organ-specific promoters may be, for example, a promoter from storage sink tissues such as seeds, potato tubers, and fruits (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), or from metabolic sink tissues such as meristems (Ita *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), a seed specific promoter such as the glutelin, prolamin, globulin, or albumin promoter from rice (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), a *Vicia faba* promoter from the legumin B4 and the unknown seed protein gene from *Vicia faba* (Conrad *et al.*, 1998, *Journal of Plant Physiology* 152: 708-711), a promoter from a seed oil body protein (Chen *et al.*, 1998, *Plant and Cell Physiology* 39: 935-941), the storage protein napA promoter from *Brassica napus*, or any other seed specific promoter known in the art, e.g., as described in WO 91/14772. Furthermore, the promoter may be a leaf specific promoter such as the *rbcS* promoter from rice or tomato (Kyozuka *et al.*, 1993, *Plant Physiology* 102: 891-1000, the chlorella virus adenine methyltransferase gene promoter (Mitra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93), or the *aldP* gene promoter from rice (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), or a wound inducible promoter such as the potato *pin2* promoter (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588).

A promoter enhancer element may also be used to achieve higher expression of the protease in the plant. For instance, the promoter enhancer element may be an intron which is placed between the promoter and the nucleotide sequence encoding a polypeptide of the

present invention. For instance, Xu et al., 1993, *supra* disclose the use of the first intron of the rice actin 1 gene to enhance expression.

Still further, the codon usage may be optimized for the plant species in question to improve expression (see Horvath et al referred to above).

5 The selectable marker gene and any other parts of the expression construct may be chosen from those available in the art.

10 The nucleic acid construct is incorporated into the plant genome according to conventional techniques known in the art, including *Agrobacterium*-mediated transformation, virus-mediated transformation, microinjection, particle bombardment, biolistic transformation, and electroporation (Gasser et al., 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto et al., 1989, *Nature* 338: 274).

15 Presently, *Agrobacterium tumefaciens*-mediated gene transfer is the method of choice for generating transgenic dicots (for a review, see Hooykaas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38). However it can also be used for transforming monocots, although other transformation methods are generally preferred for these plants. Presently, the method of choice for generating transgenic monocots is particle bombardment (microscopic gold or tungsten particles coated with the transforming DNA) of embryonic calli or developing embryos (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil et al., 1992, *Bio/Technology* 10: 667-674). An alternative method for 20 transformation of monocots is based on protoplast transformation as described by Omirulleh et al., 1993, *Plant Molecular Biology* 21: 415-428.

Following transformation, the transformants having incorporated therein the expression construct are selected and regenerated into whole plants according to methods well-known in the art.

25 The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a transgenic plant or a plant cell comprising a nucleic acid sequence encoding a polypeptide having protease activity of the present invention under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide. The invention relates to a transgenic plant, or plant part, comprising a polynucleotide as defined in 30 claim 8, or an expression vector or polynucleotide construct of the invention.

#### Animals

The present invention also relates to a transgenic, non-human animal and products or elements thereof, examples of which are body fluids such as milk and blood, organs, flesh, and

animal cells. Techniques for expressing proteins, e.g. in mammalian cells, are known in the art, see e.g. the handbook Protein Expression: A Practical Approach, Higgins and Hames (eds), Oxford University Press (1999), and the three other handbooks in this series relating to Gene Transcription, RNA processing, and Post-translational Processing. Generally speaking, to 5 prepare a transgenic animal, selected cells of a selected animal are transformed with a nucleic acid sequence encoding a polypeptide having protease activity of the present invention so as to express and produce the polypeptide. The polypeptide may be recovered from the animal, e.g. from the milk of female animals, or the polypeptide may be expressed to the benefit of the animal itself, e.g. to assist the animal's digestion. Examples of animals are mentioned below in 10 the section headed Animal Feed.

To produce a transgenic animal with a view to recovering the protease from the milk of the animal, a gene encoding the protease may be inserted into the fertilized eggs of an animal in question, e.g. by use of a transgene expression vector which comprises a suitable milk protein promoter, and the gene encoding the protease. The transgene expression vector is 15 microinjected into fertilized eggs, and preferably permanently integrated into the chromosome. Once the egg begins to grow and divide, the potential embryo is implanted into a surrogate mother, and animals carrying the transgene are identified. The resulting animal can then be multiplied by conventional breeding. The polypeptide may be purified from the animal's milk, see e.g. Meade, H.M. et al (1999): Expression of recombinant proteins in the milk of transgenic 20 animals, Gene expression systems: Using nature for the art of expression, J. M. Fernandez and J. P. Hoeffer (eds.), Academic Press.

In the alternative, in order to produce a transgenic non-human animal that carries in the genome of its somatic and/or germ cells a nucleic acid sequence including a heterologous transgene construct including a transgene encoding the protease, the transgene may be operably linked to a first regulatory sequence for salivary gland specific expression of the protease, as disclosed in WO 2000064247.

The invention relates to a transgenic, non-human animal, or products, or elements thereof, comprising a polynucleotide, or an expression vector or polynucleotide construct of the invention.

30

#### Methods of Production

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a host cell or a transgenic plant or animal under conditions

conducive for production of the polypeptide in a supernatant; and optionally (b) recovering the polypeptide.

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptides to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies, formation of a product, or disappearance of a substrate. For example, a protease assay may be used to determine the activity of the polypeptide as described herein.

The resulting polypeptide may be recovered by methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulfate precipitation), SDS-PAGE, or extraction (see, e.g., *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

#### Compositions

In a still further aspect, the present invention relates to compositions comprising a polypeptide of the present invention. The polypeptide compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. For instance, the polypeptide composition may be in the form of a granulate or a microgranulate. The polypeptide to be included in the composition may be stabilized in

accordance with methods known in the art. Examples are given below of preferred uses of the polypeptides or polypeptide compositions of the invention.

#### Animal Feed

The present invention is also directed to methods for using the polypeptides of the invention in animal feed, as well as to feed compositions and feed additives comprising the polypeptides of the invention. The term animal includes all animals, including human beings. Examples of animals are non-ruminants, and ruminants, such as cows, sheep and horses. In a particular embodiment, the animal is a non-ruminant animal. Non-ruminant animals include mono-gastric animals, e.g. pigs or swine (including, but not limited to, piglets, growing pigs, and sows); poultry such as turkeys, ducks and chicken (including but not limited to broiler chicks, layers); young calves; and fish (including but not limited to salmon, trout, tilapia, catfish and carps; and crustaceans (including but not limited to shrimps and prawns))

The term feed or feed composition means any compound, preparation, mixture, or composition suitable for, or intended for intake by an animal.

In the use according to the invention the protease can be fed to the animal before, after, or simultaneously with the diet. The latter is preferred.

In a particular embodiment, the protease, in the form in which it is added to the feed, or when being included in a feed additive, is well-defined. Well-defined means that the protease preparation is at least 50% pure as determined by Size-exclusion chromatography (see Example 12 of WO 01/58275). In other particular embodiments the protease preparation is at least 60, 70, 80, 85, 88, 90, 92, 94, or at least 95% pure as determined by this method. A well-defined protease preparation is advantageous. For instance, it is much easier to dose correctly to the feed a protease that is essentially free from interfering or contaminating other proteases. The term dose correctly refers in particular to the objective of obtaining consistent and constant results, and the capability of optimising dosage based upon the desired effect.

For the use in animal feed, however, the protease need not be that pure; it may e.g. include other enzymes, in which case it could be termed a protease preparation. The protease preparation can be (a) added directly to the feed (or used directly in a treatment process of vegetable proteins), or (b) it can be used in the production of one or more intermediate compositions such as feed additives or premixes that is subsequently added to the feed (or used in a treatment process). The degree of purity described above refers to the purity of the original protease preparation, whether used according to (a) or (b) above.

Protease preparations with purities of this order of magnitude are in particular obtainable using recombinant methods of production, whereas they are not so easily obtained and also subject to a much higher batch-to-batch variation when the protease is produced by traditional fermentation methods. Such protease preparation may of course be mixed with other enzymes.

In a particular embodiment, the protease for use according to the invention is capable of solubilising vegetable proteins. A suitable assay for determining solubilised protein is disclosed in Example 11.

The term vegetable proteins as used herein refers to any compound, composition, preparation or mixture that includes at least one protein derived from or originating from a vegetable, including modified proteins and protein-derivatives. In particular embodiments, the protein content of the vegetable proteins is at least 10, 20, 30, 40, 50, or 60% (w/w). Vegetable proteins may be derived from vegetable protein sources, such as legumes and cereals, for example materials from plants of the families Fabaceae (*Leguminosae*), Cruciferaceae, Chenopodiaceae, and Poaceae, such as soy bean meal, lupin meal and rapeseed meal. In a particular embodiment, the vegetable protein source is material from one or more plants of the family Fabaceae, e.g. soybean, lupine, pea, or bean.

In another particular embodiment, the vegetable protein source is material from one or more plants of the family Chenopodiaceae, e.g. beet, sugar beet, spinach or quinoa. Other examples of vegetable protein sources are rapeseed, and cabbage. Soybean is a preferred vegetable protein source. Other examples of vegetable protein sources are cereals such as barley, wheat, rye, oat, maize (corn), rice, and sorghum.

The treatment according to the invention of vegetable proteins with at least one protease of the invention results in an increased solubilisation of vegetable proteins. The following are examples of % solubilised protein obtainable using the proteases of the invention in a monogastric *in vitro* model: At least 102%, 103%, 104%, 105%, 106%, or at least 107%, relative to a blank. The percentage of solubilised protein is determined using the monogastric *in vitro* model of Example 11. The term solubilisation of proteins basically means bringing protein(s) into solution. Such solubilisation may be due to protease-mediated release of protein from other components of the usually complex natural compositions such as feed.

In a further particular embodiment, the protease for use according to the invention is capable of increasing the amount of digestible vegetable proteins. The following are examples of % digested or digestible protein obtainable using the proteases of the invention in a monogastric *in vitro* model: At least 104%, 105%, 106%, 107%, 108%, 109%, or at least 110%, relative to a

blank. The percentage of digested or digestible protein is determined using the *in vitro* model of Example 11.

The following are examples of % digested or digestible protein obtainable using the proteases of the invention in an aquaculture *in vitro* model: At least 103%, 104%, 105%, 106%, 107%, 108%, 109% or at least 110%, relative to a blank. The percentage of digested or digestible protein is determined using the aquaculture *in vitro* model of Example 12.

In a still further particular embodiment, the protease for use according to the invention is capable of increasing the Degree of Hydrolysis (DH) of vegetable proteins. The following are examples of Degree of Hydrolysis increase obtainable in a monogastric *in vitro* model: At least 102%, 103%, 104%, 105%, 106%, or at least 107%, relative to a blank. The DH is determined using the monogastric *in vitro* model of Example 11. The following are examples of Degree of Hydrolysis increase obtainable in an aquaculture *in vitro* model: At least 102%, 103%, 104%, 105%, 106%, or at least 107%, relative to a blank. The DH is determined using the aquaculture *in vitro* model of Example 12.

In a particular embodiment of a (pre-) treatment process of the invention, the protease(s) in question is affecting (or acting on, or exerting its solubilising influence on) the vegetable proteins or protein sources. To achieve this, the vegetable protein or protein source is typically suspended in a solvent, e.g. an aqueous solvent such as water, and the pH and temperature values are adjusted paying due regard to the characteristics of the enzyme in question. For example, the treatment may take place at a pH-value at which the activity of the actual protease is at least at least 40%, 50%, 60%, 70%, 80% or at least 90%. Likewise, for example, the treatment may take place at a temperature at which the activity of the actual protease is at least 40%, 50%, 60%, 70%, 80% or at least 90%. The above percentage activity indications are relative to the maximum activities. The enzymatic reaction is continued until the desired result is achieved, following which it may or may not be stopped by inactivating the enzyme, e.g. by a heat-treatment step.

In another particular embodiment of a treatment process of the invention, the protease action is sustained, meaning e.g. that the protease is added to the vegetable proteins or protein sources, but its solubilising influence is so to speak not switched on until later when desired, once suitable solubilising conditions are established, or once any enzyme inhibitors are inactivated, or whatever other means could have been applied to postpone the action of the enzyme.

In one embodiment the treatment is a pre-treatment of animal feed or vegetable proteins for use in animal feed.

The term improving the nutritional value of an animal feed means improving the availability and/or digestibility of the proteins, thereby leading to increased protein extraction from the diet components, higher protein yields, increased protein degradation and/or improved protein utilisation. The nutritional value of the feed is therefore increased, and the animal performance such as growth rate and/or weight gain and/or feed conversion ratio (i.e. the weight of ingested feed relative to weight gain) of the animal is/are improved.

In a particular embodiment the feed conversion ratio is increased by at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% or at least 10%. In a further particular embodiment the weight gain is increased by at least 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10% or at least 11%. These figures are relative to control experiments with no protease addition.

The feed conversion ratio (FCR) and the weight gain may be calculated as described in EEC (1986); Directive de la Commission du 9 avril 1986 fixant la méthode de calcul de la valeur énergétique des aliments composés destinés à la volaille. Journal Officiel des Communautés Européennes, L130, 53 – 54.

The protease can be added to the feed in any form, be it as a relatively pure protease, or in admixture with other components intended for addition to animal feed, i.e. in the form of animal feed additives, such as the so-called pre-mixes for animal feed.

In a further aspect the present invention relates to compositions for use in animal feed, such as animal feed, and animal feed additives, e.g. premixes.

Apart from the protease of the invention, the animal feed additives of the invention contain at least one fat-soluble vitamin, and/or at least one water soluble vitamin, and/or at least one trace mineral. The feed additive may also contain at least one macro mineral.

Further, optional, feed-additive ingredients are colouring agents, aroma compounds, stabilisers, antimicrobial peptides, including antifungal polypeptides, and/or at least one other enzyme selected from amongst phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.89); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-), phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.5); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6).

In a particular embodiment these other enzymes are well-defined (as defined above for protease preparations).

Examples of antimicrobial peptides (AMP's) are CAP18, Leucocin A, Triptocin, Protegrin-1, Thanatin, Defensin, Lactoferrin, Lactoferricin, and Ovispirin such as Novispirin (Robert Lehrer, 2000), Plectasins, and Statins, including the compounds and polypeptides

disclosed in PCT/DK02/00781 and PCT/DK02/00812, as well as variants or fragments of the above that retain antimicrobial activity.

Examples of antifungal polypeptides (AFP's) are the *Aspergillus giganteus*, and *Aspergillus niger* peptides, as well as variants and fragments thereof which retain antifungal activity, as disclosed in WO 94/01459 and WO 02/090384.

Usually fat- and water-soluble vitamins, as well as trace minerals form part of a so-called premix intended for addition to the feed, whereas macro minerals are usually separately added to the feed. A premix enriched with a protease of the invention, is an example of an animal feed additive of the invention.

In a particular embodiment, the animal feed additive of the invention is intended for being included (or prescribed as having to be included) in animal diets or feed at levels of 0.01 to 10.0%; more particularly 0.05 to 5.0%; or 0.2 to 1.0% (% meaning g additive per 100 g feed). This is so in particular for premixes.

The following are non-exclusive lists of examples of these components:

Examples of fat-soluble vitamins are vitamin A, vitamin D3, vitamin E, and vitamin K, e.g. vitamin K3.

Examples of water-soluble vitamins are vitamin B12, biotin and choline, vitamin B1, vitamin B2, vitamin B6, niacin, folic acid and pantothenate, e.g. Ca-D-pantothenate.

Examples of trace minerals are manganese, zinc, iron, copper, iodine, selenium, and cobalt.

Examples of macro minerals are calcium, phosphorus and sodium.

The nutritional requirements of these components (exemplified with poultry and piglets/pigs) are listed in Table A of WO 01/58275. Nutritional requirement means that these components should be provided in the diet in the concentrations indicated.

In the alternative, the animal feed additive of the invention comprises at least one of the individual components specified in Table A of WO 01/58275. At least one means either of, one or more of, one, or two, or three, or four and so forth up to all thirteen, or up to all fifteen individual components. More specifically, this at least one individual component is included in the additive of the invention in such an amount as to provide an in-feed-concentration within the range indicated in column four, or column five, or column six of Table A.

The present invention also relates to animal feed compositions. Animal feed compositions or diets have a relatively high content of protein. Poultry and pig diets can be characterised as indicated in Table B of WO 01/58275, columns 2-3. Fish diets can be characterised as indicated in column 4 of this Table B. Furthermore such fish diets usually have

a crude fat content of 200-310 g/kg. WO 01/58275 corresponds to US 09/779334 which is hereby incorporated by reference.

An animal feed composition according to the invention has a crude protein content of 50-800 g/kg, and furthermore comprises at least one protease as claimed herein.

Furthermore, or in the alternative (to the crude protein content indicated above), the animal feed composition of the invention has a content of metabolisable energy of 10-30 MJ/kg; and/or a content of calcium of 0.1-200 g/kg; and/or a content of available phosphorus of 0.1-200 g/kg; and/or a content of methionine of 0.1-100 g/kg; and/or a content of methionine plus cysteine of 0.1-150 g/kg; and/or a content of lysine of 0.5-50 g/kg.

In particular embodiments, the content of metabolisable energy, crude protein, calcium, phosphorus, methionine, methionine plus cysteine, and/or lysine is within any one of ranges 2, 3, 4 or 5 in Table B of WO 01/58275 (R. 2-5).

Crude protein is calculated as nitrogen (N) multiplied by a factor 6.25, i.e. Crude protein (g/kg)= N (g/kg) x 6.25. The nitrogen content is determined by the Kjeldahl method (A.O.A.C., 1984, Official Methods of Analysis 14th ed., Association of Official Analytical Chemists, Washington DC).

Metabolisable energy can be calculated on the basis of the NRC publication Nutrient requirements in swine, ninth revised edition 1988, subcommittee on swine nutrition, committee on animal nutrition, board of agriculture, national research council, National Academy Press, Washington, D.C., pp. 2-6, and the European Table of Energy Values for Poultry Feed-stuffs, Speleiderholst centre for poultry research and extension, 7361 DA Beekbergen, The Netherlands. Grafisch bedrijf Ponsen & loccijn bv, Wageningen. ISBN 90-71463-12-5.

The dietary content of calcium, available phosphorus and amino acids in complete animal diets is calculated on the basis of feed tables such as Veevoedertabel 1997, gegevens over chemische samenstelling, verterbaarheid en voederwaarde van voedermiddelen, Central Veevoederbureau, Runderweg 6, 8219 pk Lelystad. ISBN 90-72839-13-7.

In a particular embodiment, the animal feed composition of the invention contains at least one vegetable protein or protein source as defined above.

In still further particular embodiments, the animal feed composition of the invention contains 0-80% maize; and/or 0-80% sorghum; and/or 0-70% wheat; and/or 0-70% Barley; and/or 0-30% oats; and/or 0-40% soybean meal; and/or 0-10% fish meat; and/or 0-20% whey.

Animal diets can e.g. be manufactured as mash feed (non pelleted) or pelleted feed. Typically, the milled feed-stuffs are mixed and sufficient amounts of essential vitamins and minerals are added according to the specifications for the species in question. Enzymes can be

added as solid or liquid enzyme formulations. For example, a solid enzyme formulation is typically added before or during the mixing step; and a liquid enzyme preparation is typically added after the pelleting step. The enzyme may also be incorporated in a feed additive or premix.

5 The final enzyme concentration in the diet is within the range of 0.01-200 mg enzyme protein per kg diet, for example in the range of 0.5-25 mg enzyme protein per kg animal diet.

The protease should of course be applied in an effective amount, i.e. in an amount adequate for improving solubilisation and/or improving nutritional value of feed. It is at present contemplated that the enzyme is administered in one or more of the following amounts (dosage ranges): 0.01-200; 0.01-100; 0.5-100; 1-50; 5-100; 10-100; 0.05-50; or 0.10-10 – all these ranges being in mg protease enzyme protein per kg feed (ppm).

10 For determining mg enzyme protein per kg feed, the protease is purified from the feed composition, and the specific activity of the purified protease is determined using a relevant assay (see under protease activity, substrates, and assays). The protease activity of the feed composition as such is also determined using the same assay, and on the basis of these two determinations, the dosage in mg enzyme protein per kg feed is calculated.

15 The same principles apply for determining mg enzyme protein in feed additives. Of course, if a sample is available of the protease used for preparing the feed additive or the feed, the specific activity is determined from this sample (no need to purify the protease from the feed composition or the additive).

20 The present invention is further described by the following examples which should not be construed as limiting the scope of the invention.

#### Detergent Compositions

25 The protease of the invention may be added to and thus become a component of a detergent composition. The detergent composition of the invention may for example be formulated as a hand or machine laundry detergent composition including a laundry additive composition suitable for pre-treatment of stained fabrics and a rinse added fabric softener composition, or be formulated as a detergent composition for use in general household hard surface cleaning operations, or be formulated for hand or machine dishwashing operations.

30 In a specific aspect, the invention provides a detergent additive comprising the protease of the invention. The detergent additive as well as the detergent composition may comprise one or more other enzymes such as another protease, such as alkaline proteases from *Bacillus*, a

lipase, a cutinase, an amylase, a carbohydrase, a cellulase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, e.g., a laccase, and/or a peroxidase.

In general the properties of the chosen enzyme(s) should be compatible with the selected detergent, (i.e. pH-optimum, compatibility with other enzymatic and non-enzymatic ingredients, etc.), and the enzyme(s) should be present in effective amounts.

Suitable lipases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful lipases include lipases from *Humicola* (synonym *Thermomyces*), e.g. from *H. lanuginosa* (*T. lanuginosus*) as described in EP 268068 and EP 305216 or from *H. insolens* as described in WO 96/13580, a *Pseudomonas* lipase, e.g. from *P. alcaligenes* or *P. pseudoalcaligenes* (EP 218272), *P. cepacia* (EP 331376), *P. stutzeri* (GB 1,372,034), *P. fluorescens*, *Pseudomonas* sp. strain SD 705 (WO 95/06720 and WO 96/27002), *P. wisconsinensis* (WO 95/12012), a *Bacillus* lipase, e.g. from *B. subtilis* (Dartois et al. (1993), Biochimica et Biophysica Acta, 1131, 263-360), *B. stearothermophilus* (JP 64/744992) or *B. pumilus* (WO 91/16422). Other examples are lipase variants such as those described in WO 92/05249, WO 94/01541, EP 407225, EP 260105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 and WO 97/07202. Preferred commercially available lipase enzymes include Lipolase™ and Lipolase Ultra™ (Novozymes A/S).

Suitable amylases (alpha- and/or beta-) include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Amylases include, for example, alpha-amylases obtained from *Bacillus*, e.g. a special strain of *B. licheniformis*, described in more detail in GB 1,296,839. Examples of useful amylases are the variants described in WO 94/02597, WO 94/18314, WO 96/23873, and WO 97/43424, especially the variants with substitutions in one or more of the following positions: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, and 444. Commercially available amylases are Duramyl™, Termamyl™, Fungamyl™ and BAN™ (Novozymes A/S), Rapidase™ and Purastar™ (from Genencor International Inc.).

Suitable cellulases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Suitable cellulases include cellulases from the genera *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, e.g. the fungal cellulases produced from *Humicola insolens*, *Mycelopithora thermophila* and *Fusarium oxysporum* disclosed in US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 and WO 89/09259. Especially suitable cellulases are the alkaline or neutral cellulases having colour care benefits. Examples of such cellulases are cellulases described in EP 0 495257, EP 531372, WO

96/11262, WO 98/29397, WO 98/08940. Other examples are cellulase variants such as those described in WO 94/07998, EP 0 531 315, US 5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 and WO 99/01544. Commercially available cellulases include Celluzyme™, and Carezyme™ (Novozymes A/S), Clazinase™, and Puradax HA™ (Genencor International Inc.), and KAC-500(B)™ (Kao Corporation).

Suitable peroxidases/oxidases include those of plant, bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful peroxidases include peroxidases from *Coprinus*, e.g. from *C. cinereus*, and variants thereof as those described in WO 93/24818, WO 95/10602, and WO 98/15257. Commercially available peroxidases include Guardzyme™ (Novozymes).

The detergent enzyme(s) may be included in a detergent composition by adding separate additives containing one or more enzymes, or by adding a combined additive comprising all of these enzymes. A detergent additive of the invention, i.e. a separate additive or a combined additive, can be formulated e.g. as a granulate, a liquid, a slurry, etc. Preferred detergent additive formulations are granulates, in particular non-dusting granulates, liquids, in particular stabilized liquids, or slurries.

Non-dusting granulates may be produced, e.g., as disclosed in US 4,100,991 and 4,661,452 and may optionally be coated by methods known in the art. Examples of waxy coating materials are poly(ethylene oxide) products (polyethyleneglycol, PEG) with mean molar weights of 1000 to 20000; ethoxylated nonylphenols having from 16 to 50 ethylene oxide units; ethoxylated fatty alcohols in which the alcohol contains from 12 to 20 carbon atoms and in which there are 15 to 80 ethylene oxide units; fatty alcohols; fatty acids; and mono- and di- and triglycerides of fatty acids. Examples of film-forming coating materials suitable for application by fluid bed techniques are given in GB 1483591. Liquid enzyme preparations may, for instance, be stabilized by adding a polyol such as propylene glycol, a sugar or sugar alcohol, lactic acid or boric acid according to established methods. Protected enzymes may be prepared according to the method disclosed in EP 238216.

The detergent composition of the invention may be in any convenient form, e.g., a bar, a tablet, a powder, a granule, a paste or a liquid. A liquid detergent may be aqueous, typically containing up to 70 % water and 0-30 % organic solvent, or non-aqueous.

The detergent composition comprises one or more surfactants, which may be non-ionic including semi-polar and/or anionic and/or cationic and/or zwitterionic. The surfactants are typically present at a level of from 0.1% to 60% by weight.

When included therein the detergent will usually contain from about 1% to about 40% of an anionic surfactant such as linear alkylbenzenesulfonate, alpha-olefinsulfonate, alkyl sulfate (fatty alcohol sulfate), alcohol ethoxysulfate, secondary alkanesulfonate, alpha-sulfo fatty acid methyl ester, alkyl- or alkenylsuccinic acid or soap.

5 When included therein the detergent will usually contain from about 0.2% to about 40% of a non-ionic surfactant such as alcohol ethoxylate, nonylphenol ethoxylate, alkylpolyglycoside, alkyltrimethylammoniumoxide, ethoxylated fatty acid monoethanolamide, fatty acid monoethanolamide, polyhydroxy alkyl fatty acid amide, or N-acyl N-alkyl derivatives of glucosamine ("glucarnides").

10 The detergent may contain 0-65 % of a detergent builder or complexing agent such as zeolite, diphosphate, triphosphate, phosphonate, carbonate, citrate, nitrilotriacetic acid, ethylenediaminetetraacetic acid, diethylenetriaminepentaacetic acid, alkyl- or alkenylsuccinic acid, soluble silicates or layered silicates (e.g. SKS-6 from Hoechst).

15 The detergent may comprise one or more polymers. Examples are carboxymethylcellulose, poly(vinylpyrrolidone), poly (ethylene glycol), poly(vinyl alcohol), poly(vinylpyridine-N-oxide), poly(vinylimidazole), polycarboxylates such as polyacrylates, maleic/acrylic acid copolymers and lauryl methacrylate/acrylic acid copolymers.

20 The detergent may contain a bleaching system which may comprise a H<sub>2</sub>O<sub>2</sub> source such as perborate or percarbonate which may be combined with a peracid-forming bleach activator such as tetraacetylethylenediamine or nonanoyloxybenzenesulfonate. Alternatively, the bleaching system may comprise peroxyacids of e.g. the amide, imida, or sulfone type.

25 The enzyme(s) of the detergent composition of the invention may be stabilized using conventional stabilizing agents, e.g., a polyol such as propylene glycol or glycerol, a sugar or sugar alcohol, lactic acid, boric acid, or a boric acid derivative, e.g., an aromatic borate ester, or a phenyl boronic acid derivative such as 4-formylphenyl boronic acid, and the composition may be formulated as described in e.g. WO 92/19709 and WO 92/19708.

30 The detergent may also contain other conventional detergent ingredients such as e.g. fabric conditioners including clays, foam boosters, suds suppressors, anti-corrosion agents, soil-suspending agents, anti-soil redeposition agents, dyes, bactericides, optical brighteners, hydrotropes, tarnish inhibitors, or perfumes.

It is at present contemplated that in the detergent compositions any enzyme, in particular the enzyme of the invention, may be added in an amount corresponding to 0.01-100 mg of enzyme protein per liter of wash liquor, preferably 0.05-5 mg of enzyme protein per liter of wash liquor, in particular 0.1-1 mg of enzyme protein per liter of wash liquor.

The enzyme of the invention may additionally be incorporated in the detergent formulations disclosed in WO 97/07202.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

## EXAMPLES

### Materials and methods

#### 16 Strains:

*Bacillus subtilis* PL1601 (Diderichsen, B. et al. 1990. Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. *J. Bacteriol.*, 172, 4315-4321)

*Bacillus subtilis* MB1053

20 *Bacillus subtilis* PL3698-37

*Bacillus subtilis* MB1610

25 *Bacillus subtilis* PL2306. This strain is the *B. subtilis* DN1885 with disrupted *apr* and *npr* genes (Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. *J. Bacteriol.*, 172, 4315-4321) which is also disrupted in the transcriptional unit of the known *Bacillus subtilis* cellulase gene, resulting in cellulase negative cells. The disruption was performed essentially as described in (Eds. A.L. Sonenshein, J.A. Hoch and Richard Losick (1993) *Bacillus subtilis and other Gram-Positive Bacteria*, American Society for microbiology, p.618).

30

#### Procedure for isolating genomic DNA.

Harvest 1.5 ml culture and resuspend in 100 µl TEL. Leave at 37C for 30 min.

Add 500 µl thiocyanate buffer and leave at room temperature for 10 min.

Add 250 µl NH4Ac and leave at ice for 10 min.

- Add 500 µl CIA and mix.
- Transfer to a microcentrifuge and spin for 10 min. at full speed.
- Transfer supernatant to a new Eppendorf tube and add 0.54 volume cold isopropanol. Mix thoroughly.
- 5 Spin and wash the DNA pellet with 70 % EtOH.
- Resuspend the genomic DNA in 100 µl TER.

TE:	10 mM Tris-HCl, pH 7.4 1 mM EDTA, pH 8.0
TEL:	50 mg/ml Lysozyme in TE-buffer
Thiocyanate:	5M guanidium thiocyanate 100 mM EDTA 0.6 % w/v N-laurylsarcosine, sodium salt. 60 g thiocyanate, 20 ml 0.5 M EDTA, pH 8.0, 20 ml H <sub>2</sub> O dissolves at 65C. Cool down to RT and add 0.6 g N- laurylsarcosine. Add H <sub>2</sub> O to 100 ml and filter it through a 0.2 µ sterile filter.
NH <sub>4</sub> Ac:	7.5 M CH <sub>3</sub> COONH <sub>4</sub>
TER:	1 µg/ml RNase A in TE-buffer
CIA:	Chloroform/isoamyl alcohol 24:1

#### Purification of PCR bands and DNA sequencing

- 10 PCR fragment can be purified using GFX™ PCR DNA and Gel Band™ Purification Kit (Pharmacia Biotech) according to the manufacturer's instructions. The nucleotide sequences of the amplified PCR fragments are determined on an ABI PRISM™ 3700 DNA Analyzer (Perkin Elmer, USA) using 50-100 ng as template, the Taq deoxy-terminal cycle sequencing kit (Perkin Elmer, USA), fluorescent labeled terminators and 6 pmol of the sequencing primer of choice.

15

#### Media

TY: (As described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology", John Wiley and Sons, 1995).

- LB agar: (As described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995).

LB-PG agar; is LB agar supplemented with 0.5% Glucose and 0.05 M potassium phosphate, pH 7.0.

Proteolytic activity

S2A protease activity is measured using the PNA assay with succinyl-alanine-alanine-proline-phenylalanine-paranitroanilide as a substrate unless otherwise mention. The principle of the PNA assay is described in Rothgeb, T.M., Goodlander, B.D., Garrison, P.H., and Smith, L.A., Journal of the American Oil Chemists' Society, Vol. 65 (5) pp. 806-810 (1988).

10 Gene expression in *Bacillus subtilis* host

All the expressed genes in the following examples are integrated by homologous recombination on the *Bacillus subtilis* host cell genome. The genes are expressed under the control of a triple promoter system (as described in WO 99/43835), consisting of the promoters from *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), and the *Bacillus thuringiensis* *cryIIA* promoter including stabilizing sequence. The gene coding for Chloramphenicol acetyl-transferase was used as marker. (Described in eg. Diderichsen,B.; Poulsen,G.B.; Joergensen,S.T.; A useful cloning vector for *Bacillus subtilis*. Plasmid 30:312 (1993)).

20 Example 1. Construction of synthetic 10R tail-variant genes with Savinase signal

A synthetic 10R gene (10RS) encoding a S2A protease denoted 10R from *Nocardiopsis* sp. NRRL 18262 (WO 01/58276) was constructed which has the nucleotide sequence shown in SEQ ID NO: 1. This synthetic gene was fused by PCR in frame to the DNA coding for the signal peptide from SAVINASE™ (Novozymes) resulting in the coding sequence Sav-10RS which is shown in SEQ ID NO: 2. Several tail-variants of this construct were made. Compared to the Sav-10RS protease encoded by SEQ ID NO:2 the tail variant construct Sav-10RS HV0 was constructed to have 8 amino acids extra in the C-terminus: QSHVQSAP (SEQ ID NO: 3) which were encoded by the following DNA sequence extension inserted in front of the TAA stopcodon of SEQ ID NO: 2:

30 (SEQ ID NO: 4): caatcgcatgttcaatccgcgtcca

Tail variant Sav-10RS HV1 was constructed to have 4 amino acids extra in the C-terminus: QSAP (SEQ ID NO: 5), with the following DNA sequence extension inserted in front of the TAA stopcodon:

(SEQ ID NO: 6): caatoggccct

Tail variant Sav-10RS HV3 was constructed to have 2 amino acids extra in the C-terminus: QP (SEQ ID NO: 7) with the following DNA sequence extension inserted in front of the  
5 TAA stopcodon:

(SEQ ID NO: 8): caaoca

Tail variant Sav-10RS HV2 was constructed to have one amino acid extra in the C-terminus: P (SEQ ID NO: 9) with the following DNA sequence extension inserted in front of the  
10 TAA stopcodon:

(SEQ ID NO: 10): cca

The 10RS gene and the four tail-variant encoding genes were integrated by homologous recombination into the *Bacillus subtilis* MB1053 host cell genome. Chloramphenicol  
15 resistant transformants were checked for protease activity on 1% skim milk LB-PG agar plates (supplemented with 6 µg/ml chloramphenicol). Some protease positive colonies were further analyzed by DNA sequencing of the insert to ensure the correct gene DNA sequence, and five strains, each comprising one of the above constructs, were selected and denoted, respectively:  
B.*subtilis* Sav-10RS, B.*subtilis* Sav-10RS HV0, B.*subtilis* Sav-10RS HV1, B.*subtilis* Sav-10RS  
20 HV2 and B.*subtilis* Sav-10RS HV3.

#### Example 2. Fermentation yields of 10R tail-variants with Savinase signal

Fermentations for the production of the tail-variant enzymes of the invention were performed on a rotary shaking table in 500 ml baffled Erlenmeyer flasks each containing 100 ml  
25 TY supplemented with 6 mg/l chloramphenicol.

Six Erlenmeyer flasks for each of the five *B. subtilis* strains from example 1 were fermented in parallel. Two of the six Erlenmeyer flasks were incubated at 37°C (250 rpm), two at 30°C (250 rpm), and the last two at 26°C (250 rpm). A sample was taken from each shake flask at day 1, 2 and 3 and analyzed for proteolytic activity. The results are shown in tables 1-3. As it can be seen from tables 1 -3, the effect of the tails is a surprisingly high improvement on the expression level of the protease, as measured by activity in the culture broth. The effect is most pronounced at 26°C and 30°C, but is also evident at 37°C as an effect observed especially at the early stage of the fermentation.

Table 1: Relative proteolytic activities at 37°C.

	Day 1	Day 2	Day 3
Sav-10RS	1,0	1,0	1,0
Sav-10RS HV0	3,3	0,7	0,8
Sav-10RS HV1	4,7	1,3	1,2
Sav-10RS HV2	2,2	0,6	0,4
Sav-10RS HV3	5,3	1,4	1,7

Table 2: Relative proteolytic activities at 30°C.

	Day 1	Day 2	Day 3
Sav-10RS	1,0	1,0	1,0
Sav-10RS HV0	1,7	2,2	2,9
Sav-10RS HV1	4,6	3,1	4,9
Sav-10RS HV2	2,4	1,9	2,3
Sav-10RS HV3	4,8	3,0	4,4

Table 3: Relative proteolytic activities at 26°C.

	Day 1	Day 2	Day 3
Sav-10RS	1,0	1,0	1,0
Sav-10RS HV0	1,8	2,5	3,1
Sav-10RS HV1	2,5	3,6	4,3
Sav-10RS HV2	1,8	2,6	2,8
Sav-10RS HV3	2,6	3,5	4,6

### Example 3. Chromosomal integration of tail-variant genes

The following construct was used for the chromosomal integration of the tail-variant encoding genes. The coding sequence of the well-known subtilisin BPN' protease was operationally linked to a triple promoter, a marker gene was fused to this (a spectinomycin resistance gene surrounded by resolvase *rea*-sites), and pectate lyase encoding genes from *Bacillus subtilis* were fused to the construct as flanking segments comprising the 5' polynucleotide region upstream [*yfmD-ytmC-yfmB-yfmA-Pel-star*], and the 3' polynucleotide region downstream [*Pel-end-yflS-citS(start)*] of the tail-variant encoding polynucleotide, respectively. The integrational cassette was made by the joining of several different PCR fragments. After the final PCR reaction the PCR product was used for transformation of naturally competent *B. subtilis* cells. One clone denoted PL3598-37 was selected and confirmed by sequencing to contain the correct construct.

The PL3598-37 clone thus contains the following:

1. The flanking regions 100% homologous to region of the *B. subtilis* genome (appears as the upstream fragment *yfmD-yfmC-yfmB-yfmA-Pelstart* and the downstream fragment *Pel-end-yfIS-dtS(start)*).
  2. The Spectinomycin resistance gene flanked by Resolvase sites (*res*).
  3. The triple promoter region plus *CryIIA* mRNA stabilising leader sequence.
  4. The BPN' Open Reading Frame.

## Construction of triple promoter BPN' cassette

A PCR fragment comprising the integrational cassette for a BPN' library was constructed, thus operably linking a triple promoter (as described in WO 93/43835; Novozymes) to a BPN' expression cassette from a *Bacillus* strain. The triple promoter is a fusion of an optimized *Bacillus amyl-*derived promoter (as shown in WO 93/10249; Novozymes) with two promoters scBAN and cryIIA, where the first is a consensus version of the *Bacillus amyloliquefaciens* amylase BAN promoter, and the latter includes a mRNA-stabilising sequence (as described in WO 99/43835; Novozymes). Suitable primers can be derived from the publicly available sequences (Vasantha, N. et al. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. 159:811 (1984) EMBL: accession No. K02496). A KpnI and a Sall restriction site was introduced to flank the PCR fragment at each end, using the primers:

#252639 (SEQ ID NC\_11): catatccatctgggaccggcacttcgcagaatccctoctcaaaac

#251992 (SEQ ID NO: 12): cactgtccatgttttcaccatataaaaaacccatggaaacatccc

25 The KpnI and Sall restriction sites in the PCR fragment were subsequently used to  
clone the fragment into a KpnI-Sall digested PecI-Spec PCR fragment. The PecI-Spec fragment  
comprises a Spectinomycin resistance gene inserted in the middle of the *B. subtilis* Pectate lyase  
gene plus approx. 2.3 kb of upstream genomic DNA and approx. 1.7 kb downstream genomic  
DNA. The PecI-Spec fragment was produced by PCR amplification of genomic DNA from the  
30 *B. subtilis* strain MB1053, using the primers:

#179541 (SEQ ID NO: 13); gactttaaaggacggcgccggcggtttaaatgtatc

#179542 (SEQ ID NO: 14): ggcttggagacaaactcgaccacccaaaaatcgaccoctttc

### Construction of MB105S

The MB1053 *B. subtilis* strain was constructed by deletion of the pectate lyase (Pel) gene through integration of a PCR product into a wild-type *B. subtilis* type strain genome. This was achieved by a PCR amplification of genomic DNA directly downstream and upstream of the Pectate lyase gene of the *B. subtilis*.

The ends of the genomic DNA directly preceding and proceeding the *FeI* gene were elongated through primer insertion of sequences being 100% homologous to DNA sequences defined by the ends of a third PCR fragment encoding a marker gene surrounded by Resolvase (Res) sites. In this particular case the marker gene (*Spec*) conferred resistance to spectinomycin, and it was situated between two Res sites, altogether present on the plasmid pSJ3358 (described in US patent No. 5,882,886). Three different PCR fragments were initially produced.

Fragment 1: this fragment covers from the *yfmD* gene to the middle of the *Pel* gene and introduces an overhang to the Res-Spec-Res cassette at the *Pel* gene. The size of fragment 1 is 2.8 kb. The fragment was produced by a PCR amplification chromosomal DNA from the *B. subtilis* strain PL2306, using the primers:

#179541 (SEQ ID NO: 13), and

20 #179539 with overlap to #179164 Spec primer (SEQ ID NO: 16):

ccatttgtatcagaatctactggccatcctttacaaaccatgcggaaaataqtcatauqcaccc

Fragment 2: this fragment covers from the middle of the *Pel* gene to after the end of the *CfS* gene and introducing an overhang to the Res-Spec-Res cassette at the middle of the *Pel* gene. The size of fragment 2 is 2.3 kb. The fragment was produced by a PCR amplification of chromosomal DNA from the *B. subtilis* strain PL2306, using the primers:

#179542 (SEQ ID NO: 14), and

#179540 with overlap to #179153 Spec primer (SEQ ID NO: 16):

Fragment 3: this fragment contains the Spectinomycin gene surrounded by Res sites and DNA sequences in the ends overlapping with PCR fragment 1 and 2. The size of fragment 3 is 1.6 kb. Fragment 3 was produced by PCR amplification of plasmid pSJ3358, using the primers:

#179154 (SEQ ID NO: 17): gttgtaaaaacgacggccagtttgcataatgg  
#179153 (SEQ ID NO: 18): cogcg(cgacactagacacgggtaccctgatcttagatc

5 Standard conditions for the PCR reaction

For the PCR amplifications of fragment 1-3 the HiFi Expand™ PCR system (Roche) was used together with the following cycling scheme:

10           5 µl Buffer 2  
         14 µl dNTP's (1.25 mM each)  
10           2.5 µl 20 µM primer 1  
         2.5 µl 20µM primer 2  
         x µl water

To this mix 3 µl of DNA (apx. 100 ng) and 0.75 µl Enzyme mix (use hot start) is added.

Total volume is 50 µl.

15       The cycling profile is:

1 cycle of 120 sec at 94°C

Break.

10 cycles of 15 sec at 94°C, 60 sec at 60°C, 240 sec at 72°C.

20 cycles of 15 sec at 94°C, 60 sec at 60°C, (180 sec at 72°C add 20 sec pr cycle)

20       1 cycle 600 sec at 68°C.

The three PCR fragments were made and joined in later JOINING-PCR reactions. The three PCR fragments were single sharp bands and no gel purification was necessary. Only Qiagen™ PCR purification was performed prior to the following JOINING-PCR.

25       JOINING of fragment 1 + 3 (same procedure for fragment 2 + 3):

5 µl Buffer 2  
8 µl dNTP's (1.25 mM each)  
5.0 µl Fragment 3  
5.0 µl Fragment 1  
30       9.25 µl water

1 cycle of 120 sec at 94°C.

Break. Add Enzyme

10 cycles of 15 sec at 94°C, 60 sec at 60°C, 240 sec at 72°C.

## Break. Add Primers

15 cycles of 15 sec at 94°C, 60 sec at 60°C, (180 sec at 72°C add 20 sec pr cycle)  
1 cycle 600 sec at 68°C.

- 5 After the first cycle at 94°C for 120 sec there is a break, where 0.75 µl Enzyme mix is added.  
Total volume is now 45.0 µl.  
After the initial 10 cycles, there is another break in the cycling and for fragment 1+3: 2.5 µl  
(20µM #179541) and 2.5 µl (20 µM #179153) are added and for fragment 2+3: 2.5 µl (20µM  
#179542) and 2.5 µl (20 µM #179154) are added and the cycling is continued for 15 cycles  
10 more.

The PCR products were then gel purified. The size of fragment 1+3 should be 3.4 kb  
and the size of fragment 2+3 should be 3.4 kb. These two fragments were joined in a last PCR  
reaction (Expand™ long system, Roche):

- 15 5 µl Buffer 1  
14 µl dNTP's (1.25 mM each)  
5.0 µl Fragment 1+3  
5.0 µl Fragment 2+3  
20 17.75 µl water

- After the first cycle at 94°C for 120 sec there is a break, where 0.75 µl Enzyme mix is added.  
Total volume is now 45.0 µl.  
After the initial 10 cycles, there is another break in the cycling and 2.5 µl (20µM #179541) and  
25 2.5 µl (20 µM #179542) is added and the cycling is continued for 15 cycles more.  
1 cycle of 120 sec at 94°C.

## Break. Add Enzyme

10 cycles of 15 sec at 94°C, 60 sec at 60°C, 240 sec at 68°C.

- Break. Add Primers  
30 15 cycles of 15 sec at 94°C, 60 sec at 60°C, 180 sec at 68°C add 20 sec pr cycle  
1 cycle 600 sec at 68°C.

The size of the joined PCR fragment is 6.8 kb. This PCR fragment was purified using a  
Qiagen™ PCR purification kit, and 5 µl of the 50 µl eluted DNA was used to transform a

standard *B. subtilis* strain. After transformation cells were spread onto LBPG-120 $\mu$ g/ml of spectinomycin. Next day more than 1000 colonies were seen. 8 of these were checked using PCR primers from last JOINING PCR amplification yielding PCR fragment of 6.8 kb rather than the 5.2 kb expected if deletion had not occurred. Furthermore, the pectateleyase activity of the clones was checked with the Mancini Immunoassay, which showed no reactivity towards the pectateleyase activity. This taken together with the Spec resistance tells us that deletion had occurred. One such clone was selected and denoted MB1053.

Insertion of BPN' expression cassette adjacent to the res-spec-res in MB1053

The ligation mix of the digested PCR amplified triple promoter BPN' expression cassette and the KpnI-Sal digested Pcd-Spec PCR fragment was used as template in a PCR amplification using the PCR primers #179541 and #179542. This resulted in a PCR fragment of approx. 9 kb, which was used to transform *B. subtilis* PL1801 (Diderichsen, B et al. 1990. Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. J. Bacteriol., 172, 4315-4321) competent cells. The transformed cells were plated on LB-120  $\mu$ g/ml Spectinomycin agar plates with skim milk. Spectinomycin resistant colonies with large skim milk clearing zones were restreaked on Spectinomycin agar plates and analysed for the integration of the PCR fragment with PCR using the primers #179541 (SEQ ID NO: 13) and #179542 (SEQ ID NO: 14).

Appearance of a 9 kb fragment indicates that the PCR fragment has been integrated into the host cell genome. Several of these clones were sequenced to confirm integration of the expression cassette, one such clone was selected and denoted PL3598-37.

Example 4. Construction of plasmid-borne chromosomal integrational cassette

An *E. coli* plasmid-borne integrational cassette for a library may be constructed *In vivo*. An integration cassette to be used according to the method of the invention may be present on a *E. coli* plasmid (which is capable only of replication in *E. coli*, not in *B. subtilis*), the plasmid comprising:

- i) The DNA sequence encoding the Pre-Pro-domains of the subtilisin protease commonly known as Savinase, preceded by and operably linked to
- ii) a DNA sequence comprising a mRNA stabilising segment derived in this particular case from the CryIIa gene;
- iii) a marker gene (a chloramphenicol resistance gene), and

iv) genomic DNA from *Bacillus subtilis* as 5' and 3' flanking segments: The homologous 5' polynucleotide region upstream of the polynucleotide [yfmD-ytmC-yfmB-yfmA-Pel-start], and the 3' polynucleotide region downstream of the polynucleotide [Pel-end-yfIS-citS(start)], respectively.

5 The cassette was made by several cloning steps involving digestion of pUC19 plasmid and PCR fragments with appropriate restriction endonuclease sites of several different PCR fragments in the generally used plasmid pUC19. After each ligation of a PCR fragment into a plasmid, the ligation mixture was transformed into electrocompetent DH5 $\alpha$  E.coli cells that were prepared for and transformed by electroporation using a Gene Pulser™ electroporator from BIO-RAD as described by the supplier. One final plasmid construct was confirmed by sequencing to contain the correct construct as outlined above, and it was denoted pMB1508.

10

The pMB1508 plasmid thus contains the following:

- 15 i) The *CryIIIA* mRNA stabilising leader sequence including a ribosome binding sequence (RBS), operationally linked to
- ii) DNA encoding the Pre-Pro-domains of the subtilisin commonly known as Savinase, including KpnI and NotI sites for cloning;
- iii) The chloramphenicol resistance operon;
- iv) The 3' downstream flanking region [Pel-end-yfIS-citS(start)] which is 99-100%  
20 homologous to the region of the *B.subtilis*.

The four elements listed were cloned in the pUC19 vector (isolated from *E.coli* ATCC 37254; Vieira J, Messing J. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19: 259-268, 1982.) in the EcoRI and SalI sites to give pMB1508. In order for the resulting plasmid to integrate efficiently to a specified site of the *B.subtilis* genome, a new strain was established. The new strain is a derivative of *Bacillus subtilis* 168 BGSC accession number: 1A1 168 trpC2. The strain was made competent and transformed as described above. Using elements from the PL3598-37 clone described above, the new integration strain denoted MB1510 was established and characterised to contain the following elements from PL3598-37:

25

- i) The triple promoter and the mRNA stabilising element.
- ii) Flanking segments comprising the following homologous polynucleotide region [yfmD-ytmC-yfmB-yfmA-Pel-start] upstream of the triple-promoter, and the polynucleotide region [Pel-end-yfIS-citS(start)] downstream of the mRNA stabilizing element.

Thus, when using MB1510 competent cells, it is possible for the pMB1508 (or derivatives thereof) to directly integrate into the genome of MB1510 where the two flanking regions in fusion with the triple-promoter and mRNA stabilising element is located, resulting in a 5 construction where the incoming PrePro encoding DNA of pMB1508DNA has been integrated in the correct reading frame with the tripel-promoter, the mRNA stabilising element and the RBS. Thus resulting in high expression of the integrated gene from the promoter elements already present on the genome of MB1510.

Transformation efficiency was established for the *B.subtilis* strain MB1510 10 transformed with *E.coli* prepared plasmid pMB1508. For further testing of the potential of using this approach, the Savinase encoding gene of *Bacillus clausii* was PCR amplified using the two PCR primers:

Primer #317 (SEQ ID NO: 19) tggcgcaatcggtaccatgggg  
15 Primer #139 NotI (SEQ ID NO: 20) catgtgcattgcggcccgcattaacgatgtgccgatctgcgt

The resulting ~0.8 kb of the Savinase fragment and the pMB1508 plasmid are digested with KpnI and NotI, and the resulting fragments are then purified by agarose gel electrophoresis. The two fragments are ligated, and the ligation mixture is used to transform 20 competent *E. coli* cells which are then plated on LB-agar plates or placed in liquid media for growth overnight at 37°C; both types of media containing 50-100µg/ml of Ampicillin. After incubation, a plasmid prep is made of the liquid culture. The purified plasmid is used for transformation of competent cells of MB1510 (using 100-10,000 ng of plasmid per 25 transformation. The transformed cells are plated onto TY medium with 2% skimmilk and 6 µg/ml of chloramphenical for selection. After overnight incubation at 37°C clearing zones appear around those colonies wherein the integration cassette is integrated properly into the cells, indicating high Savinase expression.

This approach can also be used to make highly diverse libraries of any gene of interest expressable in *B.subtilis*, where rather than a gene encoding one enzyme, any expressable 30 polynucleotide is inserted into the plasmid pMB1508 and integrated into the MB1510 strain for subsequent screening.

#### Sequence of plasmid pMB1508 (SEQ ID NO: 21)

The plasmid pMB1508 has the following components, indicated by basepair positions:

BP 5186-395: pUC19 sequence from *E.coli* clone ATCC 37254, Vieira J, Messing J. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19: 259-268, 1982.

BP 396-1021: *EcoR* I cloning site (BP396-401) and the CryIIA mRNA stabilising element. (Described in WO 9634963-A1)

BP 1022-1412: Encodes the Pre-Pro sequence of Savinase and the *Nor*I cloning site. (Pre-Pro part described in eg. WO 9623073-A1, the *Nor*I site and the spacing between the Pre-Pro and *Nor*I was introduced by the PCR primer.

BP 1413-2512: The *Bg* II cloning site (BP1413-1418) and the Chloramphenicol acetyl-transferase operon of pDN1050 (Described in eg. Diderichsen,B.; Poulsen,G.B.; Joergensen,S.T.; A useful cloning vector for *Bacillus subtilis*. *Plasmid* 30:312 (1983)).

BP 2513-5185: The polynucleotide region [Pel-end-yfIS-citS(start)] downstream of the pelB locus of the *B.subtilis* genome. (as it appears from the publication and corresponding database of: F. Kunst, N. Ogasawara, I. Moszer, <146 other authors>, H. Yoshikawa, A. Danchin, "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*" *Nature* (1997) 390:249-256).

#### The *Bacillus subtilis* strain MB1510

MB1510 has the following specific features in and around the pelB locus:

20 i) The triple promoter and the mRNA stabilising element including a RBS (Ribosome binding sequence).  
ii) Flanking segments comprising the following homologous polynucleotide region [yfmD-ytmC-yfmB-yfmA-Pel-start] upstream of the triple promoter, and the polynucleotide region [Pel-end-yfIS-citS(start)] downstream of the mRNA stabilizing sequence.

25

#### Sequence of MB1510 genomic integration region (SEQ ID NO: 22)

BP 1-2873: corresponds to sequence of *Bacillus subtilis* genome yfmD-ytmC-yfmB-yfmA-Pel-start (as it appears from the publication and corresponding database of: F. Kunst et al. "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*" *Nature* (1997) 390:249-256).

BP 3102-4082: The triple promoter and CryIIA mRNA stabilising element plus RBS. (Described above in PL3598-37 construct).

BP 4083-5718: The polynucleotide region [Pel-end-yfIS-citS(start)] end of and downstream of the pelB locus of the *B.subtilis* genome (as it appears from the publication and

corresponding database of F. Kunst, N. Ogasawara, I. Moszer, <148 other authors>, H. Yoshikawa, A. Danchin, "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*" Nature (1997) 390:249-256).

5 Example 5. Construction of a 2 amino-acid tail-variant library

This example shows the construction of a tail-variant library. In this library two amino acids were introduced at the C-terminal of the 10R protein. Such a Tail-library may be made with the method described above using the following PCR primers in a PCR reaction using genomic DNA from *B. subtilis* 10RS as template:

10

1605 (SEQ ID NO: 23): gacggccagtgaattcgataaaagtgc

1606 (SEQ ID NO: 24): ccagatctctatnknkflgacggagtcctaacccccaaagag

wherein N = A, C, G or T; and K = T or G.

15

The resulting PCR product was digested with EcoR I and *Bgl* II and ligated into EcoR I and *Bgl* II digested pMB1503. Hereafter following the principle described above.

Chloramphenicol resistant *Bacillus subtilis* transformants were picked by a robotic colony picker from a bioassay plate and transferred into a 384 well microtiter plate (MTP) containing 0.05 X TY supplemented with 6 mg/l chloramphenicol (60µl/well). The MTPs were incubated at 26°C for 72h. After incubation each well was analyzed for proteolytic activity.

The thirty *Bacillus subtilis* transformants with highest proteolytic activity were selected for determination of the two tail amino acids in each transformant by DNA sequencing, the sequencing results are summaries in table 4 and table 5.

AA Tail	No. of transformants
TL	4
TT	4
QL	3
TP	3
LP	3
TI	2

IQ	2
QP	2
PI	2
LT	1
TQ	1
IT	1
QQ	1
PQ	1
Total	30

Table 4: column one shows the amino acid sequence of the tail, and column two shows the number of *Bacillus subtilis* transformants sequenced with that particular AA tail sequence.

Possibilities position 1	Result	Possibilities position 2	Result
K	0	K	0
R	0	R	0
T	14	T	6
I	3	I	4
Q	6	Q	5
P	3	P	8
L	4	L	7
Total	30	Total	30

Table 5: The table shows the amino acid which could be introduced by the primer used for the library construct and the actual findings by DNA sequencing of the thirty colonies isolated from screening.

#### 10 Example 6. Construction of *Bacillus subtilis* strains L2, L2 HV0, and L2 HV1

A *Bacillus subtilis* strain was made analogously with the construction of the *Bacillus subtilis* strain 10RS, with the DNA coding for the pro-form of the S2A protease from *Nocardiopsis dassonvillei* subsp. *Dassonvillei* DSM 43235, denoted L2, fused by PCR in frame

to the DNA coding for the signal peptide from SAVINASE™ (a well-known commercial protease derived from *Bacillus clausii*, available from Novozymes, Denmark), the resulting strain was denoted *Bacillus subtilis* Sav-L2.

The DNA sequence including the coding region for the pro-mature S2A protease from 5 *Nocardiopsis dassonvillei* subsp. *Dassonvillei* DSM 43235, as amplified with primers 1423 and 1475, is shown in SEQ ID NO: 25. The corresponding encoded pro-form amino acid sequence for the L2 protease is shown in SEQ ID NO: 28.

1423 (SEQ ID NO: 26): gcttttagttcatcgatcgatggctgtatcggcccccgtccacag

1475 (SEQ ID NO: 27): ggagcggatgtaaacatgcgataggccggatccgtacacccag

10

Two tail-variants of this construct were also made. Tail variant Sav-L2 HV0 was constructed to have 8 amino acids extra in the C-terminus: QSHV/QSAP (SEQ ID NO: 3), by using the DNA sequence extension inserted in front of the TAA stopcodon which is shown in SEQ ID NO: 4. Tail variant Sav-L2 HV1 was constructed to have 4 amino acids extra in the C-16 terminus: QSAP (SEQ ID NO: 5), by using the DNA sequence extension inserted in front of the TAA stopcodon which is shown in SEQ ID NO: 6. Both tail variants had the SAVINASE™ signal-peptide encoding sequence fused in frame with the pro-mature encoding sequence, just like in Sav-L2.

The Sav-L2 gene and the two tail-variants Sav-L2 HV0 and Sav-L2 HV1 were 20 integrated by homologous recombination on the *Bacillus subtilis* MB1053 host cell genome as outlined above. Chloramphenicol resistant transformants were checked for protease activity on 1% skim milk LB-PG agar plates (supplemented with 6 µg/ml chloramphenicol). Some protease positive colonies were further analyzed by DNA sequencing of the insert to confirm the correct DNA sequence, and one strain for each construct was selected and denoted *B. subtilis* Sav-L2, 25 *B. subtilis* Sav-L2 HV0, and *B. subtilis* Sav-L2 HV1, respectively.

#### Example 7. Fermentation yields of the *Bacillus* strains of example 6.

The three *B. subtilis* strains of example 6, were fermented on a rotary shaking table in 30 500 ml baffled Erlenmeyer flasks containing 100 ml TY supplemented with 6 mg/l chloramphenicol. Six Erlenmeyer flasks for each of the three *B. subtilis* strains were fermented in parallel. Two of the six Erlenmeyer flasks were incubated at 37°C (250 rpm), two at 30°C (250 rpm), and the last two at 26°C (250 rpm). A sample was taken from each shake flask at day 1, 2 and 3 and analyzed for proteolytic activity. The results are shown in tables 6-8. As it can be seen from tables 6-8, the effect of the tails also increases the expression level for the Sav-L2 protease

from *Nocardiopsis dassonvillei* subsp. *Dassonvillei* DSM 43235 when expressed in *B. subtilis*. An increase of up to 40% is observed in this experiment, but overall improvement is observed for both tail-variants at all three temperatures tested.

6 Table 6. Relative proteolytic activities at 37°C.

	Day 1	Day 2	Day 3
Sav-L2	1.0	1.0	1.0
Sav-L2 HV1	1.4	1.3	1.2
Sav-L2 HV0	1.3	1.1	1.4

Table 7. Relative proteolytic activities at 30°C.

	Day 1	Day 2	Day 3
Sav-L2	1.0	1.0	1.0
Sav-L2 HV1	1.0	1.2	1.4
Sav-L2 HV0	1.1	1.3	1.3

Table 8. Relative proteolytic activities at 26°C.

	Day 1	Day 2	Day 3
Sav-L2	1.0	1.0	1.0
Sav-L2 HV1	1.3	1.1	1.1
Sav-L2 HV0	0.2	1.1	1.1

10

#### Example 8. 10R tail-variants with heterologous pro-regions in *Bacillus*

The DNA sequence coding for the pro-region from the L2 protease from *Nocardiopsis dassonvillei* subsp. *Dassonvillei*, DSM 43235 is shown in SEQ ID NO: 29, and the corresponding amino acid sequence is shown in SEQ ID NO: 30. A *Bacillus subtilis* strain denoted L210R, similar to the *Bacillus subtilis* strain 10RS, but with the DNA coding for the pro-region of the L2 replacing the pro-region of 10RS, was made. The entire L210R protease encoding sequence incl. the pro-region of L2, is shown in SEQ ID NO: 31.

15

Two tail variants of the above construct were also made. Tail variant HV0 was constructed to have 8 amino acids extra in the C-terminus: QSHVQSAP (SEQ ID NO: 3) with the DNA shown in SEQ ID NO: 4 inserted in front of the TAA stopcodon of the encoding sequence. Tail variant HV1 was constructed to have 4 amino acids extra in the C-terminus: QSAP (SEQ ID NO: 5) with the DNA sequence shown in SEQ ID NO: 6 inserted in front of the TAA stopcodon of the encoding sequence.

20

The 10RL2 construct and the two tail variants were integrated by homologous recombination on the *Bacillus subtilis* MB1053 host cell genome. Chloramphenicol resistant transformants were checked for protease activity on 1% skim milk LB-PG agar plates

(supplemented with 6 µg/ml chloramphenicol). Some protease positive colonies were further analyzed by DNA sequencing of the insert to confirm the correct DNA sequence, and a strain for each construct was selected, and denoted *B. subtilis* L210R, *B. subtilis* L210R HV0, and *B. subtilis* L210R HV1, respectively.

5

#### Example 9. Fermentation yields of 10R tail-variants with heterologous pro-region

The six *B. subtilis* strains 10RS, 10RS HV0, 10RS HV1, L210R, L210R HV0, and L210R HV1, were fermented on a rotary shaking table in 500 ml baffled Erlenmeyer flasks containing 100 ml TY supplemented with 6 mg/l chloramphenicol. Six Erlenmeyer flasks for each of the *B. subtilis* strains were fermented in parallel. Two of the six Erlenmeyer flasks were incubated at 37°C (250 rpm), two at 30°C (250 rpm), and the last two at 26°C (250 rpm). A sample was taken from each shake flask at day 1, 2 and 3 and analyzed for proteolytic activity. The results are shown in figure 1, and in tables 9-11. As it can be seen from the results, the effect of the exchange of the proregion from 10R with the proregion from the L2 protease resulted in a surprisingly high improvement on the expression level of the 10R protease as measured by proteolytic activity in the culture broth at 37°C. The effect is most pronounced in the two tail variants.

Table 9. Relative proteolytic activities at 37°C.

	Day 1	Day 2	Day 3
10RS	1,0	1,0	1,0
10RS HV0	3,7	8,9	3,5
10RS HV1	3,9	8,5	4,3
L210R	1,9	2,3	1,6
L210R HV0	5,3	14,4	7,3
L210R HV1	9,1	20,9	7,6

20

Table 10. Relative proteolytic activities at 30°C.

	Day 1	Day 2	Day 3
10RS	1,0	1,0	1,0
10RS HV0	2,8	3,1	4,3
10RS HV1	3,6	3,6	4,9
L210R	0,6	0,4	0,9
L210R HV0	3,5	3,2	4,5
L210R HV1	3,7	3,2	4,5

Table 11. Relative proteolytic activities at 26°C.

	Day 1	Day 2	Day 3
10RS	1,0	1,0	1,0

10RS HV0	2,6	3,0	2,8
10RS HV1	3,7	3,3	3,1
L210R	0,4	0,7	0,4
L210R HV0	2,3	2,1	1,9
L210R HV1	2,2	1,7	1,7

**Example 10. Repeat of examples 1-9 with other 10R-like proteases**

Completely analogously with the above examples 1 through 9, similar experiments are carried out with the proteases of the following *Nocardiopsis* strains:

- 5 (a) *Nocardiopsis dassonvillei* NRRL 18133 as described in WO 88/03947;
- (b) *Nocardiopsis* sp. NRRL 18262 as described in WO 88/03947, the DNA and amino acid sequences of the protease derived from *Nocardiopsis* sp. NRRL 18262 are shown in DK patent application no. 1996 00013, and WO 01/58276 describes the use in animal feed of acid-stable proteases related to the protease derived from *Nocardiopsis* sp. NRRL no.
- 10 18262;
- (c) *Nocardiopsis Alba* DSM 15647; the amino acid sequence of the protease is SEQ ID NO: 33, the encoding nucleotide sequence is SEQ ID NO: 32; the gene is isolated from the genomic DNA of this strain by PCR-amplification using the two primers:  
1421 (SEQ ID NO: 34): gttcatcgatcgcatoggctgcgacoggccccctcccccagtc  
15 1604 (SEQ ID NO: 35): gggatcttatcaggcgccagggtcagacc.
- (d) *Nocardiopsis prasina* DSM 15648; the amino acid sequence of the protease is SEQ ID NO: 37, the encoding nucleotide sequence is SEQ ID NO: 36; the gene is isolated from the genomic DNA of this strain by PCR-amplification using the two primers:  
1346 (SEQ ID NO: 38): gttcatcgatcgcatoggctgcacccgacccgtccccccagtc  
20 1602 (SEQ ID NO: 39): gggatcttaggtccggagacggacggccccaggag.
- (e) *Nocardiopsis prasina* DSM 15649; the amino acid sequence of the protease is SEQ ID NO: 41, the encoding nucleotide sequence is SEQ ID NO: 40; the gene is isolated from the genomic DNA of this strain by PCR-amplification using the two primers:  
1603 (SEQ ID NO: 42): gttcatcgatcgcatoggctgcacccgaccactccccccagtc, and 1602 (SEQ ID  
25 NO: 39).

**Example 11. *In vivo* monogastric performance of a 10R-like protease from DSM 43235**

The performance of the *Nocardiopsis dassonvillei* subspecies *dassonvillei* DSM 43235 protease assayed in a monogastric *in vitro* digestion model. The performance of a purified preparation of the mature part of the protease having SEQ ID NO: 28 (prepared as described above) was tested in an *in vitro* model simulating the digestion in monogastric animals. In

particular, the protease was tested for its ability to improve solubilisation and digestion of maize-/SBM (maize-/soybean meal) proteins. In the tables below, this protease is designated "protease of the invention."

The *in vitro* system consisted of 15 flasks in which maize-/SBM substrate was initially incubated with HCl/pepsin - simulating gastric digestion - and subsequently with pancreatin - simulating intestinal digestion. 10 of the flasks were dosed with the protease at the start of the gastric phase whereas the remaining flasks served as blanks. At the end of the intestinal incubation phase samples of *in vitro* digesta were removed and analysed for solubilised and digested protein.

10

Table 12: Outline of *in vitro* digestion procedure

Components added	pH	Temperature	Time course	Simulated digestion phase
10 g maize-/SBM substrate (6:4), 41 ml HCl (0.105M)	3.0	40°C	t=0 min	Mixing
5 ml HCl (0.105M) / pepsin (3000 U/g substrate), 1 mL protease of the invention	3.0	40°C	t<30 min	Gastric digestion
16 ml H <sub>2</sub> O	3.0	40°C	t≈ 1.0 hour	Gastric digestion
7 ml NaOH (0.39M)	6.8	40°C	t≈ 1.5 hours	Intestinal digestion
5 ml NaHCO <sub>3</sub> (1M) / pancreatin (8 mg/g diet)	6.8	40°C	t≈ 2.0 hours	Intestinal digestion
Terminate incubation	7.0	40°C	t=6.0 hours	

#### Conditions

Substrate: 4 g SBM, 6 g maize (premixed)

15 pH: 3.0 stomach step/ 6.8-7.0 intestinal step

HCl: 0.105 M for 1.5 hours (i.e. 30 min HCl-substrate premixing)

pepsin: 3000 U/g diet for 1 hour

pancreatin: 8 mg/g diet for 4 hours

temperature: 40°C.

20 Replicates: 5

#### Solutions

0.39 M NaOH  
0.105 M HCl  
0.105 M HCl containing 6000 U pepsin per 5 ml  
1 M NaHCO<sub>3</sub> containing 16 mg pancreatin per ml  
5 125 mM NaAc-buffer, pH 6.0

#### Enzyme protein determinations

The amount of protease enzyme protein (in what follows, Enzyme Protein is abbreviated EP) is calculated on the basis of the A<sub>285</sub> values and the amino acid sequences (amino acid 10 compositions) using the principles outlined in S.C.Gill & P.H. von Hippel, Analytical Biochemistry 182, 319-326, (1989).

#### Experimental procedure for *in vitro* model

The experimental procedure was according to the above outline. pH was measured at 15 time 1, 2.5, and 5.5 hours. Incubations were terminated after 6 hours and samples of 30 ml were removed and placed on ice before centrifugation (10000 x g, 10 min, 4°C). Supernatants were removed and stored at -20°C.

#### Analysis

20 All samples were analysed for % degree of protein with the OPA method as well as content of solubilised and digested protein using gel filtration.

#### DH determination by the OPA-method

The Degree of Hydrolysis (DH) of protein in different samples was determined using an 25 semi-automated microtiter plate based colorimetric method (Nielsen,P.M.; Petersen,D.; Dambmann,C. Improved method for determining food protein degree of hydrolysis. J.Food Sci. 2001, 66, 642-646). The OPA reagent was prepared as follows: 7.620 g di-Na tetraborate decahydrate and 200 mg sodiumdodecyl sulphate (SDS) were dissolved in 150 ml deionized water. The reagents were completely dissolved before continuing. 160 mg o-phthal-dialdehyde 30 97% (OPA) was dissolved in 4 ml ethanol. The OPA solution was transferred quantitatively to the above-mentioned solution by rinsing with deionized water. 176 mg dithiothreitol 99% (DTT) was added to the solution that was made up to 200 ml with deionized water. A serine standard (0.9516 meqv/l) was prepared by solubilising 50 mg serine (Merck, Germany) in 600 ml deionized water.

The sample solution was prepared by diluting each sample to an absorbance (280 nm) of about 0.5. Generally, supernatants were diluted (100 ×) using an automated Tecan dilution station (Männedorf, Switzerland). All other spectrophotometer readings were performed at 340 nm using deionized water as the control. 25 µl of sample, standard and blind was dispensed into a microtiter plate. The micro-titer plate was inserted into an iEMS MF reader (Labsystems, Finland) and 200µl of OPA reagent was automatically dispensed. Plates were shaken (2 min; 700 rpm) before measuring absorbance. Finally, the DH was calculated. Eightfold determination of all samples was carried out.

10 Estimation of solubilised and digested protein

The content of solubilised protein in supernatants from *in vitro* digested samples was estimated by quantifying crude protein (CP) using gel filtration HPLC. Supernatants were thawed, filtered through 0.45 µm polycarbonate filters and diluted (1:50, v/v) with H<sub>2</sub>O. Diluted samples were chromatographed by HPLC using a Superdex Peptide PE (7.5 x 300 mm) gel filtration column (Global). The eluent used for isocratic elution was 50 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. The total volume of eluent per run was 26 ml and the flow rate was 0.4 ml/min. Elution profiles were recorded at 214 nm and the total area under the profiles was determined by integration. To estimate protein content from integrated areas, a calibration curve ( $R^2=0.9993$ ) was made from a dilution series of an *in vitro* digested reference maize/GBM sample with known total protein content. The protein determination in this reference sample was carried out using the Kjeldahl method (determination of % nitrogen; A.O.A.C. (1984) Official Methods of Analysis 14th ed., Washington DC).

The content of digested protein was estimated by integrating the chromatogram area corresponding to peptides and amino acids having a molecular mass of 1500 Dalton or below (Savole,L.; Gauthier,S.F. Dialysis Cell For The In-vitro Measurement Of Protein Digestibility. J. Food Sci. 1986, 51, 494-498; Babinszky,L.; Van,D.M.J.M.; Boer,H.; Den,H.L.A. An In-vitro Method for Prediction of The Digestible Crude Protein Content in Pig Feeds. J. Sci. Food Agr. 1990, 50, 173-178; Boisen,S.; Eggum,B.O. Critical Evaluation of In-vitro Methods for Estimating Digestibility in Simple-Stomach Animals. Nutrition Research Reviews 1991, 4, 141-162). To determine the 1500 Dalton dividing line, the gel filtration column was calibrated using cytochrome C (Boehringer, Germany), aprotinin, gastrin I, and substance P (Sigma Aldrich, USA), as molecular mass standards.

Results

The results shown in Tables 13 and 14 below indicate that the protease increased the Degree of Hydrolysis (DH), as well as soluble and digestible protein significantly.

Table 13: Degree of Hydrolysis (DH), absolute and relative values

Enzyme (dosage in mg EP/kg feed)	n	Of total protein		Relative to blank		%CV
		%DH	SD	%DH	SD	
Blank	5	26.84 <sup>a</sup>	0.69	100.0 <sup>a</sup>	2.57	
Protease of the invention (100)	5	28.21 <sup>b</sup>	0.36	105.1 <sup>b</sup>	1.25	

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%

Table 14: Solubilised and digested crude protein measured by ÄKTA HPLC

Enzyme (dosage in mg EP/kg feed)	n	Of total protein				Relative to blank			
		%dig. CP	SD	%sol.CP	SD	%dig.CP	CV%	%sol.CP	CV%
Blank	5	54.1 <sup>a</sup>	1.1	90.1 <sup>a</sup>	1.1	100.0 <sup>a</sup>	2.0	100.0 <sup>a</sup>	1.2
Protease of the invention (50)	5	57.7 <sup>b</sup>	1.1	93.2 <sup>b</sup>	1.4	106.7 <sup>b</sup>	1.9	103.4 <sup>b</sup>	1.5
(100)	5	58.9 <sup>b</sup>	0.8	94.8 <sup>b</sup>	0.9	108.9 <sup>b</sup>	1.3	105.2 <sup>b</sup>	0.9

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%

#### Example 12. *In vitro* aquaculture performance of 10R-like protease from DSM 43235

Performance of the protease from *Nocardicopsis dassonvillei* subsp. *dassonvillei* DSM 43235 in an aquaculture *in vitro* model. The protease preparation as described in Example 3 was tested in an aquaculture *in vitro* model simulating the digestion in coldwater fish. The *in vitro* system consisted of 15 flasks in which SBM substrate was initially incubated with HCl/pepsin –

simulating gastric digestion – and subsequently with pancreatin – simulating intestinal digestion. 10 of the flasks were dosed with the protease at the start of the gastric phase whereas the remaining 5 flasks served as blanks. At the end of the intestinal incubation phase samples of *in vitro* digesta were removed and analysed for solubilised and digested protein.

5

Table 15: Outline of aqua *in vitro* digestion procedure

Components added	pH	Temperature	Time course	Simulated digestion phase
10 g extruded SBM substrate, 62 mL HCl (0.155M)/pepsin (4000 U/g substrate), 1 mL of the protease of the invention	3.0	15°C	t=0 min	Gastric digestion
7 mL NaOH (1.1M)	6.8	15°C	t=6 hours	Intestinal digestion
5 mL NaHCO <sub>3</sub> (1M) / pancreatin (8 mg /g diet)	6.8	15°C	t=7 hours	Intestinal digestion
Terminate Incubation	7.0	15°C	t=24 hours	

Conditions

Substrate: 10 g extruded SBM

10 pH: 3.0 stomach step/ 6.8-7.0 intestinal step

HCl: 0.155 M for 6 hours

Pepsin: 4000 U /g diet for 6 hours

Pancreatin: 8 mg/g diet for 17 hours

Temperature: 15°C

15 Replicates: 5

Solutions

1.1 M NaOH

0.155 M HCl / pepsin (4000 U/g diet)

20 1 M NaHCO<sub>3</sub> containing 16 mg pancreatin/mL

125 mM NaAc-buffer, pH 6.0

Experimental procedure for aqua *in vitro* model

The experimental produce was according to the above outline. pH was measured at time 1, 5, 8 and 23 hours. Incubations were terminated after 24 hours and samples of 30 mL were removed and placed on ice before centrifugation (13000 x g, 10 min, 0°C). Supernatants were removed and stored at -20°C.

5

#### Analysis

All supernatants were analysed using the OPA method (% degree of hydrolysis) and by AKTA HPLC to determine solubilised and digested protein (see monogastric example).

10 Pre-treatment of *in vitro* supernatants with EASY SPE columns

Before analysis on AKTA HPLC supernatants from the *in vitro* system were pretreated using solid-phase sample purification. This was done to improve the chromatography and thereby prevent unstable elution profiles and baselines. The columns used for extraction were solid phase extraction columns (Chromabond EASY SPE Columns from Macherey-Nagel). 2 mL milliQ water was eluted through the columns by use of a vacuum chamber (vacuum 0.15 x 100 kPa). Subsequently 3 mL *in vitro* sample was dispensed onto the column and eluted (vacuum 0.1 x 100 kPa), the first ½ mL of eluted sample was thrown away and a clean tube was placed beneath the column, then the rest of the sample was eluted and saved for further dilution.

15 Results

The results shown in Tables 16 and 17 below indicate that the protease significantly increased Degree of hydrolysis and protein digestibility.

20 Table 16: Degree of Hydrolysis (DH) measured by the OPA method, absolute and relative values

Enzyme (mg EP/kg diet)	n	Of total protein			Relative to blank		
		%DH	SD	%DH	SD	%CV	
Blank	5	21.30	<sup>a</sup>	0.52	100.0	<sup>a</sup>	2.42
Protease of the invention (50)	5	21.98	<sup>b</sup>	0.22	103.2	<sup>b</sup>	1.00

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%

**Table 17:** Solubilised and digested crude protein measured by AKTA HPLC, absolute and relative values

Enzyme (mg EP/kg diet)	N	Of total protein					Relative to blank						
		%C P dig		SD	%C P sol		SD	%CP dig		%C V	%CP sol		
Blank	5	50.0	<sup>a</sup>	2.2	89.9	<sup>a</sup>	3.2	100.0	<sup>a</sup>	4.5	100.0	<sup>a</sup>	3.5
Protease of the invention (50)	5	52.3	<sup>b</sup>	1.1	91.4	<sup>a</sup>	1.5	104.8	<sup>b</sup>	2.1	101.7	<sup>a</sup>	1.6
(100)	5	53.4	<sup>b</sup>	0.4	91.6	<sup>a</sup>	1.0	107.0	<sup>b</sup>	0.7	101.9	<sup>a</sup>	1.1

Different letters within the same column indicate significant differences (1-way ANOVA, Tukey-Kramer test, P<0.05). SD = Standard Deviation. %CV = Coefficient of Variance = (SD/mean value) x 100%.

#### Example 13. Fermentation and activity of 10R tail-variants TQ and TP with Savinase signal

Two of the *B. subtilis* strains of Example 5, strain 209 with the amino acid tail-variant TQ, and strain 211 with the tail-variant TP, together with *B. subtilis* Sav-10RS, were fermented on a rotary shaking table in 600 ml baffled Erlenmeyer flasks containing 100 ml TY supplemented with 6 mg/l chloramphenicol. Twelve Erlenmeyer flasks for each of the three *B. subtilis* strains were fermented in parallel. Four of the twelve Erlenmeyer flasks were incubated at 37°C (250 rpm), four at 30°C (250 rpm), and the last four at 26°C (250 rpm). A sample was taken from each shake flask at day 1, 2 and 3 and analyzed for proteolytic activity. The results are shown in tables 18 to 20 below.

As it can be seen from tables below, the effect of the 2 amino acid tails is a surprisingly high improvement on the yield of the protease, as measured by activity in the culture broth. The effect of the 2 amino acid tails is comparable to the effect observed for Sav-10RS HV1 and Sav-10RS HV3 in Example 1.

20

Table 18: Relative proteolytic activities at 37°C.

	1	2	3
10R synt-15	1,0	1,0	1,0
209	7,0	7,0	6,0
211	7,2	7,7	4,9

Table 19: Relative proteolytic activities at 30°C.

	1	2	3
10R synt-15	1,0	1,0	1,0
209	4,5	3,6	4,9
211	4,0	4,1	5,0

Table 20: Relative proteolytic activities at 26°C.

	1	2	3
10R synt-15	1,0	1,0	1,0
209	6,4	4,3	4,0
211	3,7	4,1	4,2

5

**Example 14. Synthetic shuffled 10R-like protease tail-variants with signal**

Another synthetic tail variant 10R protease encoding gene, denoted G-MAT-22, was constructed with a signal peptide, and the 8 amino acid C-terminal tail of SEQ ID NO: 3, and 10 introduced into a *Bacillus subtilis* host for expression. A surprisingly high yield of protease was achieved (data not shown). The full coding DNA sequence of G-MAT-22 is shown in SEQ ID NO: 44, and the encoded pre-pro-protease is shown in SEQ ID NO: 45. The G-mat-22 protease is an alpha-lytic protease-like enzyme (peptidase family S1E – old notation: S2A). This protease has a higher temperature optimum (at pH 9) than the 10R protease, as shown in Figure 1.

15

**Example 15: Shuffled Pro-sequences of 10R-like Proteases**

Recombination of protease genes can be made independently of the specific sequence of the parents by synthetic shuffling as described in Ness, J.E. et al 2002 [Nature Biotechnology, Vol. 20 (12), pp. 1251-1255, 2002]. Synthetic oligonucleotides degenerated in their DNA sequence to provide the possibility of all amino acids found in the set of parent proteases are designed and the genes assembled according to the reference. The shuffling can be carried out for the full length sequence or for only part of the sequence and then later combined with the rest of the gene to give a full length sequence.

In this example the amino acid sequence for the Pro-peptide part of the parent proteases given in SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or 26 SEQ ID NO: 45 is encoded by a set of oligonucleotides and the resulting shuffled gene fragments are combined into the context of the full length protease gene, which then consists of DNA coding for the signal sequence, the (shuffled) Pro-peptide, and in this case the mature

protein of 10R protease. Examples of shuffled Pro-peptide sequences are shown in SEQ ID NO: 46 (O-2.19), SEQ ID NO: 47 (G-2.73), SEQ ID NO: 48 (G-1.43), SEQ ID NO: 49 (G-2.6), SEQ ID NO: 50 (G-2.5), SEQ ID NO: 51 (G-2.3), SEQ ID NO: 52 (G-1.4), and SEQ ID NO: 53 (G-1.2).

The complete protease encoding genes were inserted into the genome of *B.subtilis* by homologous recombination as described above, and the proteases expressed in shakeflasks using a rich media. The fermentation was carried out for 5 days at 30°C and the supernatant isolated by centrifugation prior to measuring the protease activity. As control a *B.subtilis* clone expressing the wild type protease 10R from *Nocardiopsis* sp. NRRL 18262 from an identical construction protocol was fermented under the same conditions. The protease activity was calculated and is presented in the table below relatively to the activity of the wild type 10R protease. Clearly the heterologous pro-regions provide an advantage over the native pro-region of the 10R protease.

	Ref. activity
10R	1,0
G-1.2	2,9
G-1.4	1,4
G-2.3	1,6
G-2.4	3,4
G-2.5	3,0
G-2.6	4,2
G-2.7	1,6

Table 21: Relative activity of 10R protease expressed with heterologous shuffled pro-peptides.

Example 16. *In vivo* monogastric performance of tail-variant 10R-HV1

This example describes a dose/response study with the four amino acid tail variant HV1 of the 10R protease in the monogastric in vitro model using 10, 25, 50, and 100 mg EP/kg, and using 10R protease as benchmark or control. The tail variant 10R-HV1 was constructed to have 4 amino acids extra in the C-terminus: QSAP (SEQ ID NO: 5) as described above.

20 In vitro conditions:

- |             |   |
|-------------|---|
| Substrate:  | 4 g SBM, 6 g maize (premixed)                               |
| pH:         | 3.0 stomach step/ 6.8-7.0 intestine step                    |
| HCl:        | 0.105 M for 1.5 hours (i.e. 30 min HCl-substrate premixing) |
| Pepsin:     | 3000 U /g diet for 1 hour                                   |
| Pancreatin: | 8 mg/g diet for 4 hours.                                    |
| Incubation: | 40°C.   |
| Replica:    | 5   |

Enzymes:

10R protease: FFE-2003-00047; batch PPA21400; 154 mg EP/g product.  
Freezedried 10R-HV1 FFE-2003-00077; 370 mg EP/g product

6 Solution A: 10R, 100 mg EP/kg diet:

100 mg EP/kg diet ~1 mg EP/flask  $\Rightarrow$  1 mg EP/mL  
(1 mg EP/mL \* 10 mL) / 154 mg EP/g product = 0.0649 g  
Prepare 10 mL: Dissolve 0.0649 g enzyme in 10 mL NaAc buffer.

10 Solution C: 10R-HV1, 100 mg EP/kg diet:

100 mg EP/kg diet ~1 mg EP/flask  $\Rightarrow$  1 mg EP/mL  
(1 mg EP/mL \* 20 mL) / 370 mg EP/g product = 0.05405 g  
Prepare 20 mL: Dissolve 0.0541 g enzyme in 20 mL NaAc buffer.

15 Solution D: 10R-HV1, 50 mg EP/kg diet:

50 mg EP/kg diet ~0.50 mg EP/flask via 1 ml = 0.50 mg EP/ml  
Prepare 10 mL: Dilute C 2 times: 5 ml solution C + 5 ml 125 mM NaAc-buffer

Solution E: 10R-HV1, 25 mg EP/kg diet:

20 25 mg EP/kg diet ~0.25 mg EP/flask via 1 ml = 0.25 mg EP/ml  
Prepare 12 mL: Dilute C 4 times: 3 ml solution C + 9 ml 125 mM NaAc-buffer

Solution F: 10R-HV1, 10 mg EP/kg diet:

25 mg EP/kg diet ~0.25 mg EP/flask via 1 ml = 0.25 mg EP/ml  
25 Prepare 10 mL: Dilute C 10 times: 1 ml solution C + 9 ml 125 mM NaAc-buffer

Substrates:

Premix (40% SBM / 60% maize), FFS-2002-00121

The 10 g sample contains 6 g maize and 4 g SBM giving a calculated protein content of 23.48 %  
30 of protein (~ 2.35 g/flask).

Chemicals:

4.005 M HCl, AT-1-00061/29  
4.007 M NaOH, AT-1-00002/36

Pancreatin FFE-2002-00052, 8xUSP

Pepsin FFE-2003-00048, 471 U/mg

NaOH 0.39 M:

- 5 Prepare 500 mL:  
48.97 mL 3.982 M NaOH, fill with milliQ to 500 mL.

HCl1 solution 0.105 M

- Prepare 2000 mL:  
52.43 mL of 4.005 HCl, fill with milliQ up to 2000 mL.

10 HCl2 (HCl/pepsin) solution: 0.105 M containing 30000 U pepsin/5 mL

- Prepare 250 mL:  
Take out approx. 150 mL from the HCl-solution, add 3.18 g pepsin and fill up to 250 mL with the HCl solution.

15 125 mM NaAc-buffer, pH 6.0:

- Prepared from a 2 M NaAc-buffer (KLu 04-07-2003/lab book 14169 p. 104)  
→ 12.5 mL 2 M NaAc-buffer, fill up to 200 mL with milliQ

Pancreatin dissolved in 1 M NaHCO<sub>3</sub> containing 8 mg pancreatin/g diet:

- 20 NaHCO<sub>3</sub>-pancreatin is pre made, divided into portions and frozen. Made 29-04-2003 and frozen, it is slowly thawed in refrigerator over night. The stock preparation is described in lab. book 14165 page 068.

Flow scheme:

- 25 In the Premixing phase (t=0), 10 g substrate is mixed with 41 ml HCl1; then in the gastral phase (t=30 min) 5 ml HCl-2 (HCl/pepsin) + 1 ml enzyme (or buffer) is added, and later (t=1 h) the pH is measured and 16 ml water is added; and then in the intestinal phase (t=1h 30 min) 7 ml 0.39 NaOH is added, and later (t=2h) 5 ml NaHCO<sub>3</sub>/pancreatin is added and the pH is measured again twice (t=2h 30 min & t=5h 30 min); and finally (t=6h) 30 ml suspension is sampled for centrifugation. Each supernatant is immediately and carefully removed from the centrifuge tube into glass tubes. The supernatants are split in two aliquots for further analysis. Results are shown in table 22.

Table 22: Treatment of samples in the monogastric *in vitro* model.

Sample	Enzyme Solution	Enzyme	pH	Enzyme dose /kg diet:	Pepsin U/g diet:	Pancreatin mg/g diet:
1 – 5	1 ml Buffer	Blank	3.0	0 mg EP	3000	8.0
6 – 10	1 ml Solution A	10R (FFE-2003-00047)	3.0	100 mg EP	3000	8.0
11 – 15	1 ml Solution C	10R-HV1 (PPA22873)	3.0	100 mg EP	3000	8.0
16 – 20	1 ml Solution D	10R-HV1 (PPA22873)	3.0	50 mg EP	3000	8.0
21 – 25	1 ml Solution E	10R-HV1 (PPA22873)	3.0	25 mg EP	3000	8.0
26 – 30	1 ml Solution F	10R-HV1 (PPA22873)	3.0	10 mg EP	3000	8.0

Soluble and Digestible Protein:

The changes in the levels of soluble and digestible crude protein in the soluble phase of the hydrolysates were determined using an AKTA HPLC (Superdex 30 peptide column). The results are shown in Table 23.

At a 10R-HV1 dose of 100 mg EP/kg diet, the level of Digestible protein was significantly increased by 9.8%, compared to Blank. The control 10R showed a relative improvement of 7.7%. With the lower enzyme concentrations (50, 25, and 10 mg EP/kg diet) the relative improvements of Digestible protein were 5.7%, 3.3% and 0.7%, respectively.

Table 23: HPLC results with 10R-HV1 and 10R showing the percentual changes in digestible CP and soluble CP relative to blank. Different letters on top of the bars indicate significant differences (1-ANOVA, Tukey, 95%).

Enzyme [mg EP/kg]	n	Of total protein				Relative to blank			
		%dig. CP	SD	%solCP	SD	%dig. CP	CV%	%solCP	CV%
Blank	11	54,8	1,2	83,9	1,6	100,0 <sup>a</sup>	2,3	100,0 <sup>a</sup>	1,9
10R HV1 [100]	5	60,2	0,8	88,3	1,1	109,6 <sup>b</sup>	1,3	105,3 <sup>c</sup>	1,3
10R HV1 [50]	5	58,0	0,6	86,9	0,8	105,7 <sup>cd</sup>	1,1	102,6 <sup>bc</sup>	0,9
10R HV1 [25]	5	56,7	0,8	86,3	1,0	103,3 <sup>bc</sup>	1,4	102,9 <sup>bc</sup>	1,2
10R HV1 [10]	5	55,2	1,3	84,3	1,9	100,7 <sup>ab</sup>	2,4	100,5 <sup>ab</sup>	2,2
10R [100]	5	59,1	0,8	87,1	1,9	107,7 <sup>de</sup>	1,4	103,9 <sup>bc</sup>	2,2

The original 10R [100 mg EP/kg diet] improved the level of soluble protein by about 4%. The effects of 10R-HV1 was slightly higher (5.3% relative increase) and significant. With a dose of 50 and 25 mg EP/kg diet the relative improvements were 3.6% and 2.9%, respectively and significant. With 10 mg EP/kg diet the relative improvement was 0.5%.

#### Degree of Hydrolysis:

The degree of hydrolysis (DH) was determined using the OPA method. Results are shown in Table 24.

Enzyme [mgEP/kg]	n	Of total protein		Relative to blank	
		%DH	SD	%DH	CV%
Blank	5	25,89 <sup>a</sup>	0,43	100,0 <sup>a</sup>	1,65
10R (FFE-2003-00047) [100]	5	27,19 <sup>bc</sup>	0,67	105,0 <sup>b</sup>	2,46
10R <sub>HV1</sub> [100]	5	27,89 <sup>c</sup>	0,36	107,7 <sup>c</sup>	1,29
10R <sub>HV1</sub> [50]	5	27,34 <sup>bc</sup>	0,57	105,6 <sup>b</sup>	2,08
10R <sub>HV1</sub> [25]	5	26,42 <sup>abc</sup>	0,57	102,0 <sup>ab</sup>	2,16
10R <sub>HV1</sub> [10]	5	25,52 <sup>a</sup>	0,96	98,6 <sup>a</sup>	3,76

Table 24: Degree of Hydrolysis (DH) determined by the OPA method. Absolute as well as relative values are shown. Different letters indicate significant differences (1-way ANOVA, Tukey 95 %).

Tail-variant 10R-HV1 improved DH by 7.7%, compared to Blank. With the lower doses (50 and 25 mg EP/kg diet) of the protease the improvements ranged from 5.6 - 2.0%, respectively, in line with previous findings. At the lowest dose [10 mg EP/kg diet] no effect was seen. The original 10R [100 mg EP/kg diet] showed improvements of 5% relative to Blank.

The results of the HPLC ÄKTA analysis and the DH determinations clearly show that addition of the four amino acid (SEQ ID NO: 5) long tail to 10R does not affect the performance of the 10R protease to any significant extent.

## CLAIMS

1. A secreted mature polypeptide which after maturation has protease activity, which polypeptide when expressed and before maturation comprises a heterologous pro-region, and which polypeptide:

- 5 (a) comprises an amino acid sequence which is at least 70% identical to the amino acid sequence of the mature part of the polypeptide shown in SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45;
- 10 (b) comprises an amino acid sequence which is at least 70% identical to the amino acid sequence of the mature part of the polypeptide encoded by the polynucleotide in SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 25; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 36; SEQ ID NO: 40; or SEQ ID NO: 44;
- 15 (c) comprises a mature part which is a variant of the mature part of the polypeptide having the amino acid sequence of SEQ ID NO: 28; SEQ ID NO: 33; SEQ ID NO: 37; SEQ ID NO: 41; SEQ ID NO: 43; or SEQ ID NO: 45, the segment comprising a substitution, deletion, extension, and/or insertion of one or more amino acids;
- (d) is an allelic variant of (a), (b), or (c);
- (e) is a fragment of (a), (b), (c), or (d).

2. The polypeptide according to claim 1, wherein the heterologous pro-region is derived from a protease, preferably the pro-region is derived from an S2A or S1E protease, and most preferably it is at least 70% identical to the pro-region shown in SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53.

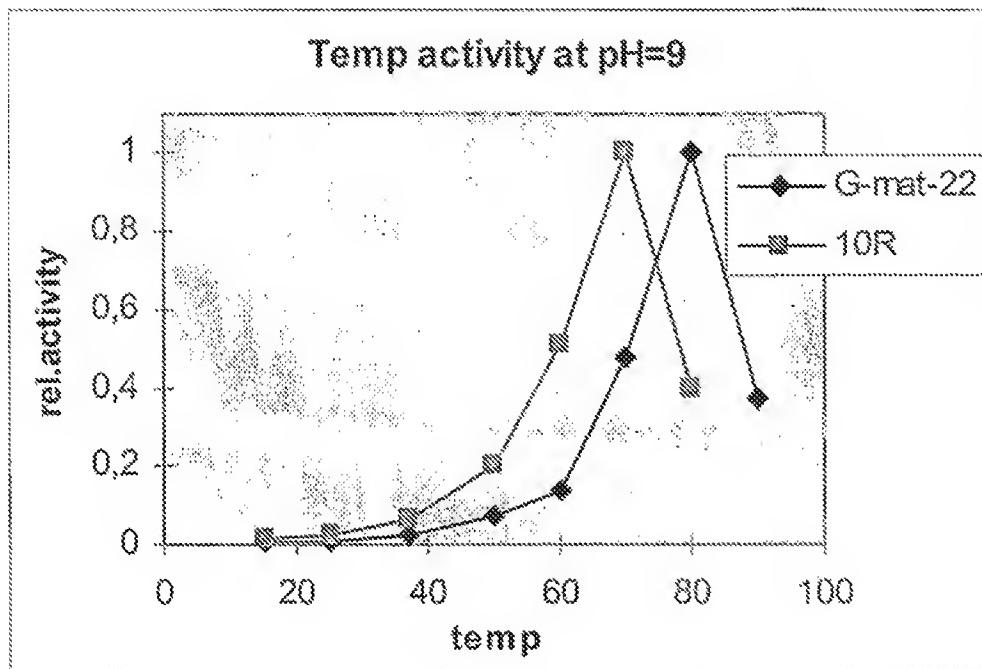
25 3. The polypeptide according to claim 1 or 2, which is a wildtype polypeptide, an artificial variant of a wildtype polypeptide said variant having one or more amino-acid(s) added to the C-terminus as compared to the wildtype, a shuffled polypeptide, or a protein-engineered polypeptide.

30 4. The polypeptide according to any of claims 1 - 3, which comprises at least three non-polar or uncharged polar amino acids within the last four amino acids of the C-terminus of the polypeptide

5. The polypeptide according to claim 3, wherein the one or more added amino acid(s) is (are) non-polar or uncharged.
6. The polypeptide according to claim 5, wherein the one or more added amino acid(s) is one or 5 more of Q, S, V, A, or P.
7. The polypeptide according to claim 3, wherein the one or more added amino acids are selected from the group consisting of: QSHVQSAP, QSAP, QP, TL, TT, QL, TP, LP, TI, IQ, QP, PI, LT, TQ, IT, QQ, and PQ.  
10
8. The polypeptide according to any of claims 1 – 7 which when expressed and before maturation comprises a heterologous secretion signal-peptide which is cleaved from the polypeptide when the polypeptide is secreted, preferably the heterologous secretion signal peptide is derived from a heterologous protease.  
15
9. The polypeptide according to claim 8, wherein the heterologous secretion signal peptide comprises an amino acid sequence having a sequence identity of at least 70% with the amino acid sequence encoded by polynucleotides 1 – 81 of SEQ ID NO: 2, or SEQ ID NO: 44.
- 20 10. An isolated polynucleotide encoding a polypeptide as defined in any of claims 1-9.
11. A recombinant expression vector or polynucleotide construct comprising a polynucleotide as defined in claim 10.
- 25 12. A recombinant host cell comprising a polynucleotide as defined in claim 10, or an expression vector or polynucleotide construct as defined in claim 11.
13. The recombinant host cell according to claim 12 which is a *Bacillus* cell.
- 30 14. A transgenic plant, or plant part, comprising a polynucleotide as defined in claim 10, or an expression vector or polynucleotide construct as defined in claim 11.

15. A transgenic, non-human animal, or products, or elements thereof, comprising a polynucleotide as defined in claim 10, or an expression vector or polynucleotide construct as defined in claim 11.
- 5 16. A method for producing a polypeptide as defined in any of claims 1 - 9, the method comprising: (a) cultivating a recombinant host cell as defined in claim 12 or 13, or a transgenic plant or animal as defined in claims 14 or 15, to produce a supernatant comprising the polypeptide, and optionally (b) recovering the polypeptide.
- 10 17. An animal feed additive comprising at least one polypeptide as defined in any of claims 1 - 9; and
  - (a) at least one fat-soluble vitamin, and/or
  - (b) at least one water-soluble vitamin, and/or
  - (c) at least one trace mineral.
- 15 18. An animal feed composition having a crude protein content of 50 to 800 g/kg and comprising at least one polypeptide as defined in any of claims 1 - 9, or at least one feed additive of claim 17.
- 20 19. A composition comprising at least one polypeptide as defined in any of claims 1 - 9, together with at least one other enzyme selected from amongst phytase (EC 3.1.3.8 or 3.1.3.26); xylanase (EC 3.2.1.8); galactanase (EC 3.2.1.89); alpha-galactosidase (EC 3.2.1.22); protease (EC 3.4.-.-); phospholipase A1 (EC 3.1.1.32); phospholipase A2 (EC 3.1.1.4); lysophospholipase (EC 3.1.1.5); phospholipase C (3.1.4.3); phospholipase D (EC 3.1.4.4); and/or beta-glucanase (EC 3.2.1.4 or EC 3.2.1.6).
- 25 20. A method for using at least one polypeptide as defined in any of claims 1 - 9, for improving the nutritional value of an animal feed, for increasing digestible and/or soluble protein in animal diets, for increasing the degree of hydrolysis of proteins in animal diets, and/or for the treatment of vegetable proteins, the method comprising including the polypeptide(s) in animal feed, and/or in a composition for use in animal feed.
- 30 21. A method for using at least one polypeptide as defined in any of claims 1 - 9, comprising including the polypeptide(s) in a detergent formulation.

1/1

**FIG. 1**

10495.204-WO.ST25.txt  
SEQUENCE LISTING

<110> Lassen, Soren Flensted  
 <120> Improved fusion proteases and methods for producing them  
 <130> 10495.204-WO  
 <160> S3  
 <170> PatentIn version 3.2  
 <230> 1  
 <211> 1062  
 <212> DNA  
 <213> Nocardiopsis sp. NRRL 18262

<220>  
 <221> misc\_feature  
 <222> (1)..(495)  
 <223> Encodes the pro-region shown in positions ~165 to ~1 of SEQ ID NO:43.

<220>  
 <221> misc\_feature  
 <222> (496)..(1059)  
 <223> Encodes the mature region shown in positions 1-188 of SEQ ID NO:43.

<400> 1  
 gctactggag cattacctca gtctccatca cctgaaggcag atgcagtatc gatgcaagaa 60  
 gcattacaac gtgatcttga tcttacatca gctgaagctg aggaattact tgctgcacaa 120  
 gatacagcct ttgaagtttg tgaagctgcc gctgaaggcag ctggtgatgc atatgggt 180  
 tcagtattcg atactgaatc actcgaacct actgtacttag tgaccgatgc agcagctgtt 240  
 gaagctgttg aagccacagg tgcaggtaca gagctcgtat cttatggtat tggatggatta 300  
 gatgagatcg tacaagaact taatycaygt gatgccgttc caggtgttgt tggatggtat 360  
 cctgatgttag caggtgatac tggatgttttta gaagttcttg aaggctctgg agctgatgtt 420  
 tctggacttt tagcagacgc aggagtcgat gcattccgcgg ttgaagtgtac cacgtcagat 480  
 cagccctgaac tctatgccga tatcatttggaa ggcctagcgt acacaatggg tggtcgtgc 540  
 agcgttaggat ttgcagccac aaatycayct ggacaacctg gcttcgtgac agctggacat 600  
 tgcggccqcg tcggcacaca gtttacttcc ggcaatggaa gaggtgtctt tgagcaaagc 660  
 gtatitcccg ggaatgtatgc tgccttcgtt agaggtacgt ccaactttac gcttacttaac 720  
 tttagtatcta gatacaacac tggccggatat gcaactgttag caggtcacaa tcaagcacct 780  
 attggctcta gcgtctgccg ctcagggtcty actacaggat ggcattgtgg aaccatcaa 840  
 octagaggct zggycytgag ctatcctgaa ggtaccgtaa cgaacatgac tcgtacgtact 900  
 gtaigtgcag aaccagggtga ctctggaggt tcatatatca gcggtacgca agcgcaaggc 960  
 gttacacctag gtggatccgg taactgttagg acaggtggca caacgttcta ccaggaagtg 1020  
 acaccgtatgg tgaactcttg gggagttaga ctccgtacat aa 1062

## 10495.204-WO.ST25.txt

<210> 2  
<211> 1143  
<212> DNA  
<213> Artificial sequence

<220>  
<223> A synthetic 10R gene (10Rsynt-15) encoding a S2A protease denoted "10R" fused by PCR in frame to the signal peptide encoding sequence of a heterologous protease, Savinase.

<400> 2						
atgaagaaac	cgttggggaa	aattgtcgca	agcaccgcac	tactcaatttc	tgttgctttt	60
agttcatcgta	tcgcacatggc	tgctactgga	gcattacctc	agtctccatc	acctgaagca	120
gatgcagtat	cgtatgcaga	agcattacaa	cgtgatcttg	atcttacatc	agctgaagct	180
gaggaattac	ttgctgcaca	agatacagcc	tttgaagttt	atgaagctgc	cgctgaagca	240
gctgggtatg	cataatggtgg	ttcagtttttc	gatactgaat	cactcgaaact	tactgtacta	300
gtgaccgatg	cagcagctgt	tgaagctgtt	gaagccacag	gtgcaggta	agagctcgta	360
tcttatggta	tttgatggatt	agatgagatc	gtacaagagc	ttaatgcagc	tgatgccgtt	420
ccaggtgttag	ttggatggta	tcctgtatgt	gcaggtgata	ctgttgtctt	agaagttctt	480
gaaggctctg	gagctgtatgt	ttctggacatt	tttagcagacg	caggagtcga	tgcattcccg	540
gttgaagtga	ccacgtcaga	tcaagccgtaa	ctctatgccg	atatcattgg	aggccctagcg	600
tacacaatgg	gtggcgctg	cagcgttagga	tttgcagcca	caaatgcagc	tggacaacct	660
ggcttcgtga	cagctggaca	ttgcggccgc	gtcggtacac	aggttactat	cgcaatgg	720
agaggtgtct	ttgagcaaag	cgtatccc	ggaaatgtatg	ctgccttcgt	tagaggtacg	780
tccaaacttta	cgcttactaa	cttagtatct	agataacaaca	ctggccgata	tgcactgtat	840
gcaggtcaca	atcaagcacc	tattggctct	agcgtctgac	gttcagggtc	gactacagga	900
tggcattgtg	gaaccattca	agcttagaggt	cagagcgtga	gttatccgt	aggtaccgt	960
acqaacatga	ctcgatcgac	tgtatgtgca	gaaccagggt	actctggagg	ttcatatatc	1020
agcggtagcgc	aagcgtcaagg	cgtracctca	ggtggatccg	gtactgttag	gacaggtggc	1080
acaacgttct	accaggaagt	gacaccgtat	gtgaactctt	ggggagttag	actccgtaca	1140
taa						1143

<210> 3  
<211> 8  
<212> PRT  
<213> Artificial sequence

<220>  
<223> C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 3

Gln Ser His Val Gln Ser Ala Pro  
1 5

<210> 4

## 10495.204-WO.ST25.txt

<211> 24  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Polynucleotide encoding a C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 4  
caatccgcatg ttcaatccgc tcca

<210> 5  
<211> 4  
<212> PRT  
<213> Artificial sequence

<220>  
<223> C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 5  
Gln Ser Ala Pro  
1

<210> 6  
<211> 12  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Polynucleotide encoding a C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 6  
caatcggctc ct

<210> 7  
<211> 2  
<212> PRT  
<213> Artificial sequence

<220>  
<223> C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 7  
Gln Pro  
1

<210> 8  
<211> 6  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Polynucleotide encoding a C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 8  
caaccca

10495.204-WO.ST25.txt

<210> 9  
<211> 1  
<212> PRT  
<213> Artificial sequence

<220>  
<223> C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 9

Pro  
1

<210> 10  
<211> 3  
<212> DNA  
<213> Artificial sequence

<220>  
<223> polynucleotide encoding a C-terminal amino acid tail expressed as fusion to protease of the invention.

<400> 10  
cca

3

<210> 11  
<211> 45  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer #2S2639

<400> 11  
catgtgcattt tgggttaccgc aacgttcgca gatgctgttg aagag

45

<210> 12  
<211> 44  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer #2S1992

<400> 12  
catgtgcattt tgggttaccgc attatggagc ggattttaaca tgcgt

44

<210> 13  
<211> 44  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer #179541

<400> 13  
gcgtttagac gcgcggccgc gagcgccgtt tggctgaatg atac

44

<210> 14  
<211> 43  
<212> DNA  
<213> Artificial sequence

## 30495.204-WO.ST25.txt

<220>  
<223> Primer #179542  
  
<400> 14  
gcgttgagac agctcgagca gggaaaaatg gaaccgcgtt rtc 43  
  
<210> 15  
<211> 64  
<212> DNA  
<213> Artificial sequence  
  
<220>  
<223> Primer #179539  
  
<400> 15  
ccatttgcac agaattcact ggccgtcggt ttacaaccat tgccggaaat agtcataaggc 60  
atcc 64  
  
<210> 16  
<211> 60  
<212> DNA  
<213> Artificial sequence  
  
<220>  
<223> Primer #179540  
  
<400> 16  
ggatccagat ctggtaccccg ggtcttagagt cgacgcggcg gttcgcgtcc ggacagcaca 60  
  
<210> 17  
<211> 37  
<212> DNA  
<213> Artificial sequence  
  
<220>  
<223> Primer #179154  
  
<400> 17  
gttgtaaaac gacggccagt gaattctgtat caaatgg 37  
  
<210> 18  
<211> 37  
<212> DNA  
<213> Artificial sequence  
  
<220>  
<223> Primer #179153  
  
<400> 18  
ccgcgtcgac actagacacg ggtacctgtat ctagatc 37  
  
<210> 19  
<211> 32  
<212> DNA  
<213> Artificial sequence  
  
<220>  
<223> Primer #317  
  
<400> 19  
tggcgcaatc ggtaccatgg gg 22

## 10495.204-WO.ST25.txt

<210>	20	
<211>	40	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	Primer #139 NotI	
<400>	20	
	catgtgcatt cggccgcatt aacgcgttgc cgcttctgcg	40
<210>	21	
<211>	7443	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	Sequence of plasmid pME1508	
<400>	21	
	tccgcgcgtt cggcgatgac ggtgaaaaacc tctgacacat gcagctcccgg gagacggta	60
	cagcttgcgtcgtc gtaaggcgat gcccggagca gacaaggcccg tcagggcgcgc tcagcggttg	120
	ttggcggtgtc tccccggctgg cttaactatgc cggcatcaga gcagattgtt ctggatgtgc	180
	accatatgcg gtgtgaaata ccgcacagat gcgttaaggag aaaataccgc atcaggcgcc	240
	attcgccatt caggctgcgc aactgttggg aaggcgatc ggtgcgggcc tcttcgttat	300
	tacgcccagct ggcgaaaagggg ggatgtgctg caaggcgatt aagttggta acgcccagggt	360
	tttcccagtc acgacgttgtt aaaaacgacgg ccagtgaatt cgataaaaagt gcttttttttg	420
	ttgcaatttga agaatttatta atgtttaagct taattttaaaga taatatcttt gaattttaac	480
	gccccctcaaa agttagaaact acaaaaaaagg aatacgttat atagaaaatat gtttgaacct	540
	tccrtcagatt acaaataat tccggacggac tctacactcaa atgttttatctt aactatagaa	600
	tgacataccaa gcacaaacctt gaaaatttga aaatataact accaatgaac ttgttcatgt	660
	gaatttacgc tggatattttaat trtctcaatt caatataaa tatgccaata cattgttaca	720
	agttagaaatt aagacacccct tgatagccctt actataccta acatgtatgtt gtatttaaatg	780
	aatatgtaaa tatattttatg ataagaagcg actttttttt aatcattaca tatttttctt	840
	ttggaaatgtat taagatttcca atagaatagt gtataaatttta tttatcttgc aaggagggat	900
	gcctaaaaac gaagaaacatt aaaaacatata atttgcaccg tctaatggat ttatgaaaaaa	960
	tcattttatac agtttggaaaa ttatgttatta tggagctctg aaaaaaaaaaggaa gaggataaaag	1020
	aatgaagaaaa ccgttgggaa aaattgtcgc aagcacccgca ctactcattt ctgttgcttt	1080
	tgttttcatcg atcgcatcgg ctgctgaaga agcaaaaaggaa aaatattttaa ttggctttaa	1140
	tgagcaggaa gctgtcagtgc agtttgcata acaagtagag gcaaatgacg aggtcgccat	1200
	tctctctgtgag gaagagggaaatg tccggaaatttgc attgttttgcata gaattttgaaa cgttttgcgtt	1260
	tttatccgtt gagtttaagcc cagaagatgt ggacgcgcgtt gaaactcgatc cagcgatttc	1320
	ttatatttgcata gaggatgcag aagtaacgcac aatggcgcaa tccggatccat ggggtatatc	1380

## 10495.204-WO.5T25.txt

aacgcgttaa	tccgcggata	tatagcggcc	gcagatctgg	gaccaataat	aatgactaga	1440
gaagaaaagaa	tgaagattgt	tcatgaaatt	zaggAACGAA	tattggataa	agtggatAT	1500
ttttaaaata	tatattttagt	ttacagtaat	attgactttt	aaaaaaggat	tgattctaat	1560
gaagaaaAGCA	gacaAGTAAG	cctcctaaat	tcactttaga	taaaaattta	ggaggcata	1620
caaataact	tttaataaaat	tgatTTAGAC	aatttGGAAGA	gaaaAGAGAT	atTTAATCAT	1680
tatttgaacc	aacaaacgac	tttttagtata	accacAGAAA	ttgatattag	tgttttatac	1740
cgaaaacataa	aacaAGAAGG	atataaaattt	taccctgcAT	ttatTTTCTT	agtgacaagg	1800
gtgataaaact	caaatacAGC	tttttagaact	ggttacaata	gcgacGGAGA	gttaggttat	1860
tggataagt	tagGCCACT	ttatACAATT	tttgatGGTG	tatCTAAAAC	attCTCTGGT	1920
atTTGGACTC	ctgttaAGAA	tgacttcaaa	gagttttatg	atTTTACCT	ttctgtatgt	1980
gagaaaatata	atggGTCGGG	gaaattgttt	cccAAACAC	ctatacctGA	aatgttttt	2040
tctctttcca	ttatTTCCATG	gactTCATTT	actgggttta	acttAAATAT	caataataat	2100
agtaattacc	ttcttACCCAT	tattACAGCA	ggaaaattca	taataaagg	taattcaata	2160
tatTTACCAG	tatCTTACA	ggtacatcat	tctgtttgt	atggttatca	tycaggattg	2220
tttatgaact	ctatTCAGGA	attgtcagat	aggcctaATG	actggttttt	ataatATGAG	2280
ataatGCCGA	ctgtactttt	taCAGTCGGT	tttctAACGA	tacattaata	ggtacgaaaa	2340
agcaactttt	tttgcgcTTA	aaaccAGTC	taccaataac	ttaagggtaa	ctagCTCgc	2400
cggaaAGAGC	gaaaatGCCT	cacatttGTG	ccacctaAAA	aggAGCGATT	tacatATGAG	2460
ttatgcagtt	tgtAGAATGC	aaaaAGTGA	atCAGCTGG	ctaaaAGGGG	ccgcAGAGTA	2520
gaatggaaaa	ggggatCGGA	aaacaAGTAT	atAGGAGGAG	acctatTTAT	ggctTCAGAA	2580
aaAGACGCG	gaaaACAGTC	AGCAGTAAAG	tttgcTTCCAT	tgcttattac	tgtcGCTGTG	2640
ggactaatca	tctggTTTAT	tcccgctCCG	tccggactTG	aacctaAAGC	ttggcattttG	2700
tttgcgatTT	ttgtcgcAAC	atttatCGGC	tttatCTCCA	agccCTTGCC	aatgggtgca	2760
attgcaattt	ttgcattGGC	gttactGCA	ctaactGGAA	cactatcaat	tgaggatACA	2820
ttaAGCGGAT	tcggGAATAA	gaccatttgg	tttatcgTTA	tcgcattCTT	tatTTCCCGG	2880
ggatttatca	aaaccGGTCT	cggTgcgaga	attTCGTATG	tattcgTTCA	gaaattcyga	2940
aaaaaaaaACCC	ttggactttc	ttattcactG	ctatTCAGTG	atTTAATACT	ttcacCTGCT	3000
attCCAAGTC	atacGGCGCG	tgcaggAGGC	attatTTTC	ctattatCAG	atcatTTATCC	3060
gaaACATTG	gtcaAGCCCC	ggCAAATGGA	acAGAGAGAA	aaatCGGTGC	attCTTATTA	3120
aaaaACCGGTT	ttcAGGGGAA	tctgatcaca	tctgCTATGT	tcctgacAGC	gtggcGGCG	3180
aaaaCCGCTG	ttGCCAAGCT	ggccCATGAT	gtcgcAGGGG	tggacttaAC	atggacaAGC	3240
tggGCAATTG	ccgcGATTYT	accGGGACTT	gtaaGCTTAA	tcatCACGCC	gtttgtGATT	3300
tacaaaACTGT	atCCGCCGGA	aatCAAAGAA	acaccGGATG	cggcgAAAAT	cgcaACAGAA	3360
aaactgaaAG	aaatGGGACC	gttcaAAAAAA	tcggAGCTTT	ccatggTTAT	cgtgtttCTT	3420

## 10495.204-WO.ST25.txt

tttgtgtcttg	tgcgtgtggat	ttttggcgcc	agcttcaaca	tcgacgctac	cacaacccga	3480
ttgatcggtt	tggccgttct	cttattatca	caagttctga	cttgggatga	tatcaagaaa	3540
gaacacgggcg	cttgggatac	gctcaactgg	tttgcggcgc	ttgtcatgt	cgcctaacttc	3600
ttgaatgaat	taggcattgtt	gtcttgggtc	agtaatgcc	tgaaatcatc	cgtatcagg	3660
ttctcttgga	ttgtggcatt	catcatrta	attgttgtgt	attattactc	tcactatccc	3720
tttgcaagt	cyacagccca	catcagtgcg	atgtatrcag	catttttggc	tgtcgctgt	3780
gcagcgggcg	caccgcccgt	tttagcagcg	ctgagcctcg	cgttcatcag	caacctgttc	3840
gggtcaacga	ctcaactacgg	ttctggagcg	gctccggct	tcttcggagc	aggctacatc	3900
ccgcaaggca	aatggtggtc	catcgattt	atccctgtcga	ttgttcatat	catcgatgg	3960
cttgtgatcg	gcggattatg	gtggaaagta	ctaggaatat	gttagaaaga	aaaaggcaga	4020
cgcggctgtc	ctttttttat	tttcaactcct	tcgtaagaaa	atggatttgc	aaaaatgaga	4080
aaatccccgt	tgaaaaatgg	targatctag	gtagaaazgg	cggtctgggtc	tgtgtgaaa	4140
aagcgggtcc	atttttccct	gcaaaacaaa	ataatggggc	tgattgcggc	tctgtgtgtc	4200
tttgtcattt	gtgtgtgtac	cattacgtt	gccgttcagc	atacacaggg	agaacggaga	4260
caggcagagc	agctggcggt	tcaaacggcg	agaaccattt	cctatatgcc	gccgggtaaa	4320
gagctcattt	agagaaaaaga	cggacatgcg	gctcagacgc	aagagggtcat	tgaacaaaatg	4380
aaagaacaga	ctgggtcggt	tgccatttat	gttttgaacg	aaaaaggaga	cattcgctgc	4440
gcctctggaa	aaagcggatt	aaayaaactg	gagcgcagca	gagaaaatttt	gtttggcggt	4500
tcgcattttt	ctgaaacaaa	agcggatgg	cgaagagtga	tcaagagggag	cycgccccatt	4560
ataaaaagaac	agaagggata	cagccaagtg	atcggcagcg	tgtctgttga	ttttttgtca	4620
acggagacag	agcaaaagcat	caaaaagcat	tttggaaaattt	tgagtgtgtat	tgtgtgtgtt	4680
gtactgtgc	tcggatttat	tggcgccgccc	gtgctggcga	aaagcatcag	aaaggatacg	4740
cicgggcttg	aaaccgcattt	gatcgcggct	ctatatcgt	agaggaacgc	aatgccttttc	4800
gcgatttcgag	aaggattat	tgccaccaat	cgtgaaggcg	tgcgtaccat	gtgtgtgtgt	4860
tcggcggccg	agatgtgaa	gctgccccag	cctgtgtatcc	atcttcctat	agatgacgtc	4920
atgcggggag	cagggctgtat	gtctgtgttt	gaaaaaggag	aatgtgtgtc	gaaccaggaa	4980
gtaaagcgtca	acgatcaagt	gttttattatc	aatacggaaag	tgtgtgtgt	aggcggggcg	5040
gcgtatggaa	ttgtgtgtcg	tttcagggag	aaaacagagc	tgaagaagct	gatcgcacaca	5100
ttgacagagg	tttcgcataata	ttcagaggat	ctcaggccgc	agactcatgt	atttttcaaat	5160
aagctttatg	cgatttttagg	gctgcgtcga	cctgcaggca	tgcgtatgt	gcgtatcat	5220
ggtcatatgt	gttttctgtg	tgaaaattgtt	atccgttcac	aattccacac	aacatcagg	5280
ccggaagcat	aaagtgtaaa	gcctggggtg	cctaatgagt	gagcttaactc	acattaatrg	5340
cgttgcgttc	actgccccgt	ttccagtcgg	gaaacctgtc	gtgcccagctg	cattaaatgaa	5400
tcggccaaacg	cgcggggaga	ggcgggttgc	gtattggcg	ctcttcggct	tcttcgtca	5460

## 10495.204-WO.5T25.txt

ctgactcgct	ggcgtcggtc	gttcggctgc	ggcgagcggt	atcagctcac	tcaaaggcyy	5520
taatacggtt	atccacagaa	tcagggata	acgcaggaaa	gaacatgtga	gcaaaaggcc	5580
agcaaaaggc	caggaacctgt	aaaaaggccg	cgttgctggc	gtttttccat	aggctccgcc	5640
ccccctgacga	gcacacaaa	aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	5700
tataaaagata	ccagggcgrrt	ccccctggaa	gtccctcggt	gctgctctcc	gttccgacccc	5760
tgccgcttac	cggtatcccg	tccgccttcc	tcccttcggg	aagcgtggcg	ctttctcata	5820
gctcacgctg	taggtatctc	agttcggrgt	aggtcgttcg	ctccaagctg	ggctgtgtgc	5880
acgaacccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcg	tttgagtcc	5940
acccggtaag	acacgactta	tcgcccactgg	cagcagccac	tggtaacagg	attagcagag	6000
cggaggtatgt	aggcggtgt	acagagttt	tgaagtggtg	gcctaactac	ggctacacta	6060
gaaggacagt	atttggtatac	tgcgctctgc	tgaagccagt	taccttcgg	aaaagagttg	6120
gtagctcttg	atccggcaaaa	caaaccaccc	ctggtagcgg	tttttttttt	grrrycaagc	6180
agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgcatttt	tctacgggggt	6240
ctgacgctca	gtggAACGAA	aactcacgtt	aagggatrrt	ggtcatgaga	ttatcaaaaa	6300
ggatcttcac	ctagatcc	ttaaattaaa	aatgaagt	ttaaatcaatc	taasgtatat	6360
atgagtaaac	ttggcttgac	agttaccaat	gcttaatcag	tgaggcacct	atctcagcga	6420
tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	actacgatac	6480
gggagggctt	accatctggc	cccagtgcgt	caatgatacc	gcgagacccc	cgctcaccgg	6540
cacccggat	tttacagcaata	aaccagccag	ccggaaaggggc	cgagcgcaga	agtggccctg	6600
caactttatc	cgcctccatc	cagtctat	attgttgccg	ggaagctaga	graagtagtt	6660
cgccagttaa	tagtttgcc	aacgttgttg	ccattgtac	aggcatcg	gtgtcacgct	6720
cgtcgtttgg	tatggcttca	ttagctcg	gttcccaacq	atcaaggcg	gttacatgt	6780
ccccccatgtt	gtgcaaaaaaaa	gcgggttagt	ccttcggtcc	tccgatcg	gtcagaagta	6840
agttggccgc	agtgttatca	ctcafggta	tggcagcact	gcataattct	tttactgtca	6900
tgccatccgt	aagatgtttt	tctgtgactg	gtgagta	actc	ttctgagaat	6960
agtgttatgc	gcgaccgagt	tgctcttgcc	cgccgtcaat	acgggataat	accgcgccac	7020
atagcagaac	tttaaaaatgt	ctcatcat	gaaaacgttc	ttcggggcga	aaactctcaa	7080
ggatcttacc	gtctttgaga	tccagttcga	tgtaacccac	tcgtgcaccc	aactgtatctt	7140
cagcatctt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	caaaatgccg	7200
caaaaaagg	aataaggcg	acacggaaat	gttgaatact	catacttttc	tttttcaat	7260
attattgaag	cattttatcg	ggttattgtc	tcatgagcgg	atacatat	aatgtat	7320
agaaaaat	acaataggg	gttcccgcc	cattttcccg	aaaagtgc	cctgacgtct	7380
aagaaaccat	tattatcatg	acattaaacct	ataaaaatag	gcgtatc	aggcccttc	7440
gtc						7443

## 10495.204-WO.ST25.txt

<210> 22  
 <211> 5718  
 <212> DNA  
 <213> Artificial sequence  
  
 <220>  
 <223> Sequence of MS1510 genomic integration region  
  
 <400> 22  
 gagcgccgtt tggctgaatg atacaacagt ctcacttcct tactgcgtct ggttgcaaaa 60  
 acgaagaagc aaggattccc ctcgcttctc atttgtccta tttattatac acrrttttaa 120  
 gcacatcttt ggcgttggt tcacttagact tgatgcctct gaatcttgc caagtgtcac 180  
 ggtccgcacatc atagacttgt ccatttttca ccgctttagag arrrttccag agcgggttcg 240  
 ttttccactc atctacaatg gttttgcctt cgrtgctga gatgaacaaa atatcaggat 300  
 cgatrrttgct caattgctca aggctgacct cttgataggc gttatctgac ttccacagcgt 360  
 gtgtaaaagcc tagcatttttta aagatrrtctc cgtcatagga tgatgatgtt tgaagctgga 420  
 aggaatccgc tcttgcaacg ccgagaacga tggcggtt ttcatctttc ggaagttcgg 480  
 cttrtttagatc gttgatgact tttttgtgtt cggcaagctt ttcttttccct tcatctttctt 540  
 tatttaatgc tttagcaatg gtcgtaaaagc tgctgatcgt ttctgtcatat gtcgttccac 600  
 ggctttttaa ttcaatcgtc gggcgatrrr ttttcagctg ttataaaatg ttttatggc 660  
 gctcagcgtc agcgatgatt aaatcaggct tcaaggaact gatgacctca agattgggtt 720  
 cgctgcgtgt gcctacagat gtgtaatcaa tggagctgcc gacaagcttt ttaatcatat 780  
 cttrttttgtt gtcatctgctg atgccccacgg gcgtaatgcc gagattgtga acggcatcca 840  
 agaatgaaag ctcaagcaca accacccgcg taggtgtgcc gcttactgtc gttttccctt 900  
 cttcgtcatg gatcactctg gaatcccttag actcgctttt gccgcttccg ttgttattct 960  
 ggcttgcgtt acagccggat acaatgaggc aggcgagcaa taaaacactc atgatggcaa 1020  
 tcaacttgtt agaataggtg cycatgtcat ttttcctttt ttctgattttta gtaatgagaa 1080  
 tcattatcac atgtaaacact ataatacgat ggcttatcat gtcaatatrrr ttttagtaaa 1140  
 gaaagctgcg ttttactgc ttctctcatga aagcatcatc agacacaaaat aagtggatgt 1200  
 cagcgttacc gtgtcttcga gacaaaaaacg catggcggtt ggcttttagag gtttgcaca 1260  
 tatcagcagt gacataaggc aggagagtg ttagataacc ggacaatttc tttctatttt 1320  
 catctgttag tgcaaattca atgtcgccga tattcatgtat aatcgagaaa acaaagtcga 1380  
 tatcgatatg aaaatgttcc tcggcaaaaaa ccgcaagctc gtgaattccct ggtgaacatc 1440  
 cggcacgcgtt atggaaaaatc tgtttgacta aatcactcac aatccaagca ttgttattgt 1500  
 gttctgggtt aagatattgc attagacata cctccigctc gtacggataa aggcagcgtt 1560  
 ttctgggtgtt gtcgtccgtg cagcggcttc tccttaattt tgatrrttct gaaaataggt 1620  
 ccogttccctt tcactttacc atggacggas aacaaatagc tactaccatt ctcctgttt 1680  
 ttctcttcaa tggtctggaa tctgttttagt gtacagacga tcgggtatgt aagaaatata 1740

## 10495.204-WO.ST25.txt

gaaaacatga	aggaggaata	tcgacatgaa	accagttgtt	aaagagtata	caaatgacga	1800
acagctcatg	aaagatgttag	aggaatttgc	gaaaatgggt	gttgcgaaag	aggatgtata	1860
cgtcttagct	cacgacgatg	acagaacgga	acgcctggct	gacaacacga	acgccaacac	1920
gatcggagcc	aaagaaaacag	gtttcaagca	cgcgggtggga	aatatcttca	ataaaaaaagg	1980
agacgagctc	cycataaaaa	ttcacgaat	cggtttttct	gaagatgaag	ccgctcaatt	2040
tgaaaaaacgc	ttagatgaag	gaaaagtgt	tctctttgt	acagataacg	aaaaaagtgaa	2100
agcttggc	taaagcaagg	aaaaaacc	aaggccaat	tcggcctttt	ggtttttttt	2160
cggtcttgc	ggtgggattt	tgcagaatgc	cgcaatagga	tagcggaaaca	tttcggtrc	2220
tgaatgtccc	tcaatttgct	atttatatttt	tgtgatasat	tggataaaaa	tctcacaaaa	2280
tagaaaaatgg	gggtacatag	tggatgaaaa	aaatgtatgtt	agctacggct	ttgttttttag	2340
gattgactcc	agctggcg	aacgcagctg	atttaggcca	ccagacgtt	ggatccaatg	2400
atggctgggg	cgcgtactcg	accggcacga	caggcggatc	aaaagcatcc	tccctcaatg	2460
tgtataccgt	cagcaacaga	aaccagctg	tctcggcatt	agggaaaggaa	acgaacacaa	2520
cgcacaaaaat	catttatatac	aagggacacg	ttgacatgaa	cgtggatgac	aatctgaagc	2580
cgtttggcct	aatgactat	aaagatccgg	agtatgattt	ggacaaat	ttgaaagcct	2640
atgatcctag	cacatggggc	aaaaaagagc	cgtcgaaac	acaagaagaa	gcgagagcac	2700
gctctcagaa	aaaccaaaaa	gcacgggtca	tggtgat	ccctgcaaa	acgacgatcg	2760
tcggttcagg	gactaacgct	aaagtcgtgg	gaggaaactt	ccaaatcaag	egtgataacg	2820
tcattattcg	caacattgaa	ttccaggatg	cctatgacta	ttttccgc	ttgttgtaaa	2880
acgacggcca	gtgaattctg	atcaaattgt	tcagtggag	cgaagcgaac	acttgat	2940
ttaattttct	atctttata	ggtcatttata	gtataactt	ttgtcctata	aactatttag	3000
cagcataata	gatrtatttga	ataggtcatt	taagttgagc	atattagagg	aggaaaatct	3060
tggagaaaata	tttgaagaac	ccgagatcta	gtcaggtad	cgcaacgttc	gcagatgctg	3120
ttgaagagat	tataaaaaag	ctgasagcaa	aggctatca	attggtaact	gtatctcagc	3180
ttgaagaagt	gaagaagcag	agaggctatt	gaataaaatg	gtagaaagcg	ccatatcg	3240
gtttttctt	tggaaagaaaa	tatagggaaa	atggtacttg	ttaaaaattc	ggaatattt	3300
tacaatatac	tatgtatcac	attgaaagg	ggggcctgct	gtccagactg	tccgctgtgt	3360
aaaaataagg	aataaagggg	ggtgtacatt	attttactga	tatgtataat	ataatttgta	3420
taagaaaaatg	gagggccct	cgaaacgtaa	gtgaaacct	tagataaaag	tgctttttt	3480
gttgcaattg	aagaatttatt	atgtttaagc	ttaattaaag	ataatatctt	tgaatttgtaa	3540
cgccttcaa	aagtaagaac	tacaaaaaaa	gaatacgtt	tatagaaata	tgtttgaacc	3600
ttcttcagat	tacaatata	ttcggacyga	ctctacctca	aatgcttattc	taactataga	3660
atgacataca	agcacaacct	tggaaat	aaatataac	taccaatgaa	cttgttcatg	3720
tgaattatcg	ctgtat	ttttaa	ttttctcaat	tcaatata	atatgcaat	3780

## 10495.204-WO.ST25.txt

aagttagaaat	taagacacccc	ttgatagccr	tactatacct	aacatgtatgt	agtattaaat	3840
gaatatgtaa	atataatttat	gataagaagc	gacttattta	taatcattac	atattttct	3900
attggaatga	ttaagattcc	aatagaatag	tgtataaatt	atttatcttg	aaaggaggga	3960
tgcctaaaaa	cgaagaacat	taaaaacata	tatttcacc	gtctaattgga	tttatgaaaa	4020
atcattttat	cagtttggaaa	attatgtattt	atggagctct	aaaaaaaagg	agaggataaa	4080
gagaaaaaggg	gatcgaaaaa	caagtatata	ggaggagacc	tatattatggc	ttcagaaaaa	4140
gacgcagggaa	aacagtccgc	agtaaagccr	gttccattgc	tttattactgt	cgttgtggga	4200
ctaatcatct	ggtttattcc	cgctccgtcc	ggacttgaac	ctaaagcttg	gcatttgttt	4260
gcgatttttg	tcgcaacaat	tatcggtttt	atctccaagc	ccttgc当地	gggtgcaatt	4320
gcaatttttg	cattggcggt	tactgcacta	actggaaacac	tatcaattga	ggatacatta	4380
agcggattcg	ggaataayac	catttggctt	atcggtatcg	cattctttat	ttccccggga	4440
tttatcaaaa	ccyytctcgg	tgcgagaatt	tcgtatgtat	tcgttcagaa	attcggaaaa	4500
aaaacccttg	gactttctta	ttcactgc当地	ttcagtgtatt	taataactttc	acctgttatt	4560
ccaagtaata	cggcgcgtgc	aggaggcatt	atatttccta	ttatcagatc	arratccgaa	4620
scattcggat	caagccccgc	aaatggaaca	gagagaaaaa	tcgggtgc当地	cttattaaaa	4680
accggttttc	agggaaatct	gatcacatct	gctatgttcc	tgacagcgt	ggcggcgaac	4740
ccgctgattt	ccaaagctggc	ccatgatgtc	gcaggggtgg	acttaacatg	gacaagctgg	4800
gcatttgc当地	cgattgttacc	gggacttgtt	agcttaatca	tcacgccc当地	tgtgattttac	4860
aaactgttac	cgccggaaat	caaagaaaaca	ccggatgc当地	cgaaaaatcgc	aacagaaaaa	4920
ctgaaaagaaa	tgggaccgtt	aaaaaaaaatcg	gagctttcca	tggttatcg	gtttcttttg	4980
gtgcttgc当地	tgtggatttt	tggcggc当地	ttcaacatcg	acgttaccac	aaccgcattt	5040
atcggtttgg	ccgttcttctt	attatcacaa	gttctgactt	gggatgatata	caagaaaagaa	5100
cagggc当地	gggatacgtt	cacttggtrr	gcggc当地	tcatgctc当地	caacttcttg	5160
aatgaattag	gcattgggttc	ttggttc当地	aatgccatga	aatcatccgt	atcagggttc	5220
tcttggattt	tggcatttcat	catttttaatt	gttgtgtatt	attactctca	ctatattttt	5280
ycaaagtgc当地	cagccccacat	cagtgc当地	tattcagcat	ttttggctgt	cgtcggtggca	5340
gcggggc当地	cgccgctttt	agcagcgctg	agcctc当地	tcatcagcaa	cctgttc当地	5400
tcaacgactc	actacggttt	tggagc当地	ccggcttctt	tggagcagg	ctacatccc当地	5460
caaggcaaat	ggtggccat	cgatatttac	ctgtcgattt	ttcatatcat	cgtatggctt	5520
gtgatcggcg	gattatggty	gaaagtacta	ggaatatgg	aguauaaaaa	aggcagacgc	5580
ggtctgc当地	tttttattttt	cactccctcg	taagaaaaatg	gattttgaaa	aatgagaaaa	5640
tccctgtga	aaaatggtat	gatcttaggt	gaaaggacgg	ctgggtgctgt	ggtgaaaaag	5700
cggttccatt	tttccctg					5718

## 10495.204-WO.ST25.txt

<210> 23		
<211> 27		
<212> DNA		
<213> Artificial sequence		
<220>		
<223> Primer 1605		
<400> 23		
gacggccagt gaattcgata aaagtgc		27
<210> 24		
<211> 42		
<212> DNA		
<213> Artificial sequence		
<220>		
<223> Primer 1606		
<220>		
<221> misc_feature		
<222> (13)..(13)		
<223> n is a, c, g, or t		
<220>		
<221> misc_feature		
<222> (16)..(16)		
<223> n is a, c, g, or t		
<400> 24		
ccagatctt atnktnktgt acggagtcata actccccaaag ag		42
<210> 25		
<211> 1112		
<212> DNA		
<213> Nocardiopsis dassonvillei DSM 43235		
<400> 25		
gttttttagtt catcgatcgc atcggtctgtt ccggccccccg tccccccagac ccccgtegccc		60
gacgacagcg ccgcacgcat gaccgaggcg ctcaagcgcg acctcgacacct cacctcgcc		120
gaggccgagg agcrtctctc ggccgcaggaa gcccgcattcg agaccgcgc cgaggccacc		180
gaggccgcgg gcgaggccta cggcggctca ctgttcgaca ccgagaccct cgaactcacc		240
gtgctggtca ccgcgcctc cggcgatcgag gcggtegagg ccaccggagc ccaggccacc		300
gtcgtctccc acggcaccga gggcttgacc gaggtcggtgg aggacctcaa cggcgcgcgag		360
gttcccggaga gcgtccctcggtt cttgttaccccg gacgtggaga ggcgcacccgt cgtggatcgag		420
gtgctggagg gtcggcgtcgcc gcccgtctcg ccgcgcgcgg tgtggactcc		480
tccctcggtcc ggggtggagga ggcgcaggag gccccgcagg tctacgcgcga catcatcgcc		540
ggccctggcct actacatggg cggccgtcg tccgtcggtt tcgcgcgcgc caacagcgcc		600
ggtcagcccg gtttcgtcac cgccggccac tgcggcaccgg tcggcaccgg cgtgaccatc		660
ggcaacggca cggcacctt ccagaactcg gtctttcccg gcaacgcgc cgccttcgtc		720
cgcggcacct ccaacttcac cctgaccaac ctggtctcgcc gctacaactc cggcggtac		780
cagtccgtga cgggttaccag ccaggcccccg gcccgtcg cctgtgtcccg ctccggctcc		840

## 10495.204-WO.ST25.txt

accaccggct ggcaactgcgg caccatccag gccccaaacc agaccgtgcg ctacccgcag 900  
 ggcaccgtct actcgctcac ccgcaccaac gtgtgcgccc agcccgccga ctccggcgg 960  
 tcgttcatct cggctcyca ggcggaggc gtcacccctccg gcggctccgg caactgctcc 1020  
 gtcggcggca cgaccctacta ccaggaggc accccgatga tcaactccctg gggtgtcagg 1080  
 atccggacct aatcgatgt tcaatccgct cc 1112

<210> 26  
 <211> 48  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Primer 1423

<400> 26  
 gcttttagtt catcgatcgc atcggctgct ccggcccccgg tcccccaag 48

<210> 27  
 <211> 45  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Primer 1475

<400> 27  
 ggagcggatt gaacatgcga tttaggtccgg atcctgacac cccag 45

<210> 28  
 <211> 354  
 <212> PRT  
 <213> Nocardiopsis dassonvillei DSM 43235

<220>  
 <221> PROPEP  
 <222> (1)..(166)

<220>  
 <221> mat\_peptide  
 <222> (167)..(354)

<400> 28

Ala Pro Ala Pro Val Pro Gln Thr Pro Val Ala Asp Asp Ser Ala  
 -165 -160 -155

Ala Ser Met Thr Glu Ala Leu Lys Arg Asp Leu Asp Leu Thr Ser  
 -150 -145 -140

Ala Glu Ala Glu Glu Leu Leu Ser Ala Gln Glu Ala Ala Ile Glu  
 -135 -130 -125

Thr Asp Ala Glu Ala Thr Glu Ala Ala Gly Glu Ala Tyr Gly Gly  
 -120 -115 -110

10495\_204-WO\_ST25.txt

Ser	Leu	Phe	Asp	Thr	Glu	Thr	Leu	Glu	Leu	Thr	Val	Leu	Val	Thr	Asp
-105					-100					-95					
Ala	Ser	Ala	Val	Glu	Ala	Val	Glu	Ala	Thr	Gly	Ala	Gln	Ala	Thr	Val
-90				-85					-80				-75		
val	Ser	His	Gly	Thr	Glu	Gly	Leu	Thr	Glu	Val	Val	Glu	Asp	Leu	Asn
-70				-65					-60				-55		
Gly	Ala	Glu	Val	Pro	Glu	Ser	Val	Leu	Gly	Trp	Tyr	Pro	Asp	Val	Glu
-55				-50				-45							
Ser	Asp	Thr	Val	Val	Val	Glu	Val	Leu	Glu	Gly	Ser	Asp	Ala	Asp	Val
-40				-35				-30				-25			
Ala	Ala	Leu	Leu	Ala	Asp	Ala	Gly	Val	Asp	Ser	Ser	Ser	Val	Arg	Val
-25				-20				-15							
Glu	Glu	Ala	Glu	Glu	Ala	Pro	Gln	Val	Tyr	Ala	Asp	Ile	Ile	Gly	Gly
-10				-5			-1	1						5	
Leu	Ala	Tyr	Tyr	Met	Gly	Gly	Arg	Cys	Ser	Val	Gly	Phe	Ala	Ala	Thr
10				15						20					
Asn	Ser	Ala	Gly	Gln	Pro	Gly	Phe	Val	Thr	Ala	Gly	His	Cys	Gly	Thr
25				30				35							
val	Gly	Thr	Gly	Val	Thr	Ile	Gly	Asn	Gly	Thr	Gly	Thr	Phe	Gln	Asn
40				45				50							
Ser	Val	Phe	Pro	Gly	Asn	Asp	Ala	Ala	Phe	Val	Arg	Gly	Thr	Ser	Asn
55				60				65				70			
Phe	Thr	Leu	Thr	Asn	Leu	Val	Ser	Arg	Tyr	Asn	Ser	Gly	Gly	Tyr	Gln
75				80				85							
Ser	Val	Thr	Gly	Thr	Ser	Gln	Ala	Pro	Ala	Gly	Ser	Ala	Val	Cys	Arg
90				95				100							
Ser	Gly	Ser	Thr	Thr	Gly	Trp	His	Cys	Gly	Thr	Ile	Gln	Ala	Arg	Asn
105				110				115							
Gln	Thr	Val	Arg	Tyr	Pro	Gln	Gly	Thr	Val	Tyr	Ser	Leu	Thr	Arg	Thr
120				125				130							
Asn	Val	Cys	Ala	Glu	Pro	Gly	Asp	Ser	Gly	Gly	Ser	Phe	Ile	Ser	Gly
135				140				145						150	
Ser	Gln	Ala	Gln	Gly	Val	Thr	Ser	Gly	Gly	Ser	Gly	Asn	Cys	Ser	Val
155				160				165							

10495\_204-WO\_ST25.txt

Gly Gly Thr Thr Tyr Tyr Gln Glu Val Thr Pro Met Ile Asn Ser Trp  
 170 175 180

Gly Val Arg Ile Arg Thr  
 185

&lt;210&gt; 29

&lt;211&gt; 498

&lt;212&gt; DNA

&lt;213&gt; Nocardiopsis dassonvillei DSM 43235

&lt;400&gt; 29

gctccggccc ccgtccccca gaccccccgtc gccgacgaca gcgcgcgcag catgaccgag 60  
 gcgcgtcaaggc gcgcacctcg cctcacctcg gccgagggccg aggagcttct ctgcggcgac 120  
 gaagccgcga tcgagaccga cgccgaggcc accgaggccg cgggcgaggc ctacggcgcc 180  
 tcactyttcg acaccgagac cctcgaactc accgtgctgg tcaccgacgc cccggccgtc 240  
 gaggcgggtcg aggccaccgg agcccaggcc accgtcgctc cccacggcac cgaggccctg 300  
 accgaggtcg tggaggaccc caacggcgcc gaggttccccg agagcgctcct cggctggta 360  
 cggacgtgg agagcgacac cgtcgtggc gaggtgctgg agggctccga cggccacgtc 420  
 gccgcccgtc tcgcccacgc cgggtgtggac tcctccctgg tccgggtggg ggaggccgag 480  
 gaggccccgc aggtctac 498

&lt;210&gt; 30

&lt;211&gt; 166

&lt;212&gt; PRT

&lt;213&gt; Nocardiopsis dassonvillei DSM 43235

&lt;400&gt; 30

Ala Pro Ala Pro Val Pro Gln Thr Pro Val Ala Asp Asp Ser Ala Ala  
 1 5 10 15

Ser Met Thr Glu Ala Leu Lys Arg Asp Leu Asp Leu Thr Ser Ala Glu  
 20 25 30

Ala Glu Glu Leu Leu Ser Ala Gln Glu Ala Ala Ile Glu Thr Asp Ala  
 35 40 45

Glu Ala Thr Glu Ala Ala Gly Glu Ala Tyr Gly Gly Ser Leu Phe Asp  
 50 55 60

Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp Ala Ser Ala Val  
 65 70 75 80

Glu Ala Val Glu Ala Thr Gly Ala Gln Ala Thr Val Val Ser His Gly  
 85 90 95

Thr Glu Gly Leu Thr Glu Val Val Glu Asp Leu Asn Gly Ala Glu Val  
 100 105 110

10495\_204-WO\_ST25.txt

Pro	Glu	Ser	Val	Leu	Gly	Trp	Tyr	Pro	Asp	Val	Glu	Ser	Asp	Thr	Val
115								120			125				
val	val	glu	val	leu	gly	ser	asp	ala	asp	val	ala	ala	leu	leu	
130						135					140				
ala	asp	ala	gly	val	asp	ser	ser	ser	val	arg	val	glu	glu	ala	glu
145					150				155					160	
glu	ala	pro	gln	val	tyr										
					165										

<210> 31  
<211> 1146  
<212> DNA  
<213> Artificial sequence

<220>  
<223> The DNA sequence coding for the pro-region of SEQ ID NO: 29 fused in frame to A1918L2 protease tail-variant encoding gene; whole construct; 10R(proA1918L2).

<400> 31	atgaagaaac	cgttgggaa	aattgtcgca	agcacccgac	tactcatttc	tgttgctttt	60
	agttcatcgta	tcgcacatggc	tgctccggcc	cccgtcccc	agaccccccgt	cgcgcacgac	120
	agcgcgcgcca	gcatgaccga	ggcgctcaag	cgcgcacccgt	acctcaccc	ggccgaggcc	180
	gaggagcttc	tctcgccgca	ggaagccgcc	atcgagaccg	acgcccgggc	caccgaggcc	240
	gcgggcgagg	cctacggccgg	crcactgttc	gacaccgaga	ccctcyaact	caccgtgctg	300
	gtcaccgacg	cctccgcgt	cgaggccgtc	gaggccacccg	gagcccaggc	caccgtcgtc	360
	tcccacggca	ccgagggcct	gaccgaggcc	gtgggggacc	tcaacggcgc	cgaggttccc	420
	gagagcgtcc	tccggctggta	cccggacgtg	gagagcgtaca	ccgtcggtgg	cgaggtgctg	480
	gagggctccg	acgcccacgt	cgccgcctcg	ctcgccgacg	ccggtgtgg	cicctccctcg	540
	gtccgggtgg	aggaggccga	ggaggccccg	caggcttatg	ccgatatatcat	tggaggcccta	600
	grgtacacaa	tgggtggtcg	ctgcagcgta	ggatttgcag	ccacaaaatgc	ayctggacaa	660
	cctggcttcg	tgacagctgg	acattgcggc	cgcgtcggt	cacaggttac	tatcgcaat	720
	ggaagaggtg	tctttgagca	aagcgatattt	cccggaaatg	atgctgcctt	cgttagaggt	780
	acgtccaaact	ttacgcttac	taaccttagta	tctagataca	acactggccgg	atatgcaact	840
	gttagcaggc	acaatcaagc	acccratggc	tctagcgtct	gccgctcagg	gtcgactaca	900
	ggatggcatt	gtggaaaccat	tcaagctaga	ggtcagagcg	ttagctatcc	tgaaggtacc	960
	gtaaacgaaca	tgactcgta	gactgtatgt	gcagaaccag	gtgactctgg	aggttcatat	1020
	atcagcggt	cgcaagcgca	aggcggtacc	tcagggtggat	ccggtaactg	taggacaggt	1080
	ggcacaacgt	tctaccagga	agtgcacaccg	atggtgaact	cttggggagt	tagactccgt	1140
	acataaa						1146

## 30495.204-WO.ST25.txt

<210> 32  
<211> 1068  
<212> DNA

<213> Nocardiopsis Alba DSM 15647

<400> 32

gcgaccggcc	ccctccccca	gtcccccacc	ccggatgaag	ccgaggccat	caccatggtc	60
gaggccctcc	agcgcgaccc	cggccgttcc	ccctctcagg	ccgacgagct	cctcgaggcc	120
caggccgagt	ccttcgagat	cgatcgaggcc	gccaccgtgg	ccgcagccga	ctcctacggc	180
ggctccatct	tcgacaccga	cagcctcacc	ctgaccgtcc	tggtcaccga	cgcctccgccc	240
gtcgaggccg	tcgaggccgc	cggcgccgag	gccaaaggtagg	tctcgacargg	catggagggc	300
ctggaggaga	tcgtcgccga	cctqaacycg	gccgacgctc	agcccgccgt	cgtgggctgg	360
taccccgaca	tccactccga	cacggtcgtc	ctcgagggtcc	tcgaggggctc	cggtgccgac	420
gtggactccc	tgctcgccga	cggccgtgtg	gacaccgccc	acgtcaaggt	ggagagcacc	480
accgagcage	ccgagctgtt	cggcgacatc	atcggtcgcc	tcgcctacac	catgggtggg	540
cgctgctcg	tcggcttcgc	ggccaccaac	gcctccggcc	agcccggttt	cgtcaccgccc	600
ggccactgcg	gcaccgtcgg	caccccggtc	agcatcggtc	acggccaggg	cgtcttcgag	660
cgttccgtct	tccccggcaa	cgactccgccc	ttcggtccgg	gcacctcgaa	tttcacccctg	720
accaacctgg	tcagccgcta	caacaccgtt	ggttacgcga	ccgtctccgg	ctcctcgcag	780
gcggcgatcg	gttcgcagat	ctgcccgttcc	ggctccacca	ccggctggca	tcgcggcacc	840
gtccaggccc	gcggccagac	ggtgagctac	ccccagggtca	ccgtgcagaa	cctgacccgc	900
accaacgtct	gcgcgcgagcc	cggtgactcc	ggcggtctcc	tcatctccgg	cagccaggcc	960
cagggcgtca	cccccgggtgg	ctccggcaac	tgctccttcg	gtggcaccac	ctactaccag	1020
gaggtcaacc	cgatgctgag	cagctggggt	ctgaccctgc	gcacctga		1068

<210> 33  
<211> 355  
<212> PRT  
<213> Nocardiopsis Alba DSM 15647

<220>  
<221> PROPEP  
<222> (1)..(167)

<220>  
<221> mat\_peptide  
<222> (168)..(355)

<400> 33

Ala Thr Gly	Pro Leu Pro Gln Ser	Pro Thr Pro Asp Glu	Ala Glu
-165	-160	-155	

Ala Thr Thr Met Val Glu Ala Leu Gln Arg Asp Leu Gly	Leu Ser		
-150	-145	-140	

Pro Ser Gln Ala Asp Glu Leu Leu Glu Ala Gln Ala Glu	Ser Phe
	Page 18

30495.204-WO-ST25.txt  
 -135                    -130                    -125

Glu Ile Asp Glu Ala Ala Thr Ala Ala Ala Ala Asp Ser Tyr Gly  
 -120                    -115                    -110

Gly Ser Ile Phe Asp Thr Asp Ser Leu Thr Leu Thr Val Leu Val Thr  
 -105                    -100                    -95

Asp Ala Ser Ala Val Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys  
 -90                    -85                    -80

Val Val Ser His Gly Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu  
 -75                    -70                    -65                    -60

Asn Ala Ala Asp Ala Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile  
 -55                    -50                    -45

His Ser Asp Thr Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp  
 -40                    -35                    -30

Val Asp Ser Leu Leu Ala Asp Ala Gly Val Asp Thr Ala Asp Val Lys  
 -25                    -20                    -15

Val Glu Ser Thr Thr Glu Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly  
 -10                    -5                    -1                    1                    5

Gly Leu Ala Tyr Thr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala  
 10                    15                    20

Thr Asn Ala Ser Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly  
 25                    30                    35

Thr Val Gly Thr Pro Val Ser Ile Gly Asn Gly Gln Gly Val Phe Glu  
 40                    45                    50

Arg Ser Val Phe Pro Gly Asn Asp Ser Ala Phe Val Arg Gly Thr Ser  
 55                    60                    65

Asn Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr Asn Thr Gly Gly Tyr  
 70                    75                    80                    85

Ala Thr Val Ser Gly Ser Ser Gln Ala Ala Ile Gly Ser Gln Ile Cys  
 90                    95                    100

Arg Ser Gly Ser Thr Thr Gly Trp His Cys Gly Thr Val Gln Ala Arg  
 105                    110                    115

Gly Gln Thr Val Ser Tyr Pro Gln Gly Thr Val Gln Asn Leu Thr Arg  
 120                    125                    130

Thr Asn Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Phe Ile Ser

135                    140                    145  
 150                    155                    160  
 165

Gly Ser Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Ser  
 150                    155                    160                    165

Phe Gly Gly Thr Thr Tyr Tyr Gln Glu Val Asn Pro Met Leu Ser Ser  
 170                    175                    180

Trp Gly Leu Thr Leu Arg Thr  
 185

<210> 34  
<211> 43  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer 1421

<400> 34  
gttcatcgat cgcacatggct gcgaccggcc ccctccccca gtc

43

<210> 35  
<211> 31  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer 1604

<400> 35  
gcggatccta tcaggtgcgc agggtcagac c

31

<210> 36  
<211> 1062  
<212> DNA  
<213> Nocardiopsis prasina DSM 15648

<400> 36  
gccacggac cgctccccc gtcacccacc cggaggccg acggcgatcc catgcaggag  
ggcgctccagc gcgacctcg cctgaccccg cttgaggccg atgaactgtt ggccgcccag  
gacaccgcct tcgaggtcga cgaggcccg gccgcggccg cggggacgc ctacggcgcc  
tccgtttcg acaccgagac cctggaaactg accgtctgg tcaccgacgc cgcctcggtc  
gaggctgtgg aggccacccgg cgccggtaacc gaactcgatc cctacggcat cgagggcctc  
gacgagatca tctaggatct caacgcgcgc gacgcggatcc cggcggttgt cggctggtag  
ccggacgtgg cgggtgacac cgtcgccctg gaggtctgg agggttccgg agccgacgtg  
agccggctgc tcgcccacgc cggcggtggac gcctcgcccg tcgaggtgac cagcagtgcg  
cagccccggc tctacgcccga catcatcgcc ggtctggccct acaccatggg cggccgtgt  
tcgggtcgat tcgcccacca aacgcgcgc ggtcagcccg gattcgatcac cggccgtcac  
tgtggcccg tgccacccca ggtgagcatc ggcaacggcc agggcgttt cggcgttcc  
atcttcccg gcaacgacgc cggccgtcac cggccgtcac gctgaccaac

60

120

180

240

300

360

420

480

540

600

660

720

## 10495.204-WO.5T25.txt

ctggtcagcc gctacaacac cggcggttac gccaccgtcg ccggccacaa ccaggcgccc	780
atcggctcct ccgtctgccg ctccggctcc accaccggct ggcactgcgg caccatccag	840
gccccgcggcc agtcggtgag ctaccccgag ggcaccgtca ccaacatgac ccggaccacc	900
gtgtgcgcgg agcccgccga ctccggcggc tcctacatct ccggcaacca ggcccagggc	960
gtcacctccg gcggtcccg caactgcccgc accggcgaaa ccaccttcta ccaggaggtc	1020
accccatgg tgaactcccg gggcgccgt ctccggacct aa	1062

&lt;210&gt; 37

&lt;211&gt; 353

&lt;212&gt; PRT

&lt;213&gt; Nocardiopsis prasina DSM 15648

&lt;220&gt;

&lt;221&gt; PROPEP

&lt;222&gt; (1)..(165)

&lt;220&gt;

&lt;221&gt; mat\_peptide

&lt;222&gt; (166)..(353)

&lt;400&gt; 37

Ala	Thr	Gly	Pro	Leu	Pro	Gln	Ser	Pro	Thr	Pro	Glu	Ala	Asp	Ala
-165														

-160														

-155														

Val	Ser	Met	Gln	Glu	Ala	Leu	Gln	Arg	Asp	Leu	Gly	Leu	Thr	Pro
-150														

-145														

-140														

Leu	Glu	Ala	Asp	Glu	Leu	Leu	Ala	Ala	Gln	Asp	Thr	Ala	Phe	Glu
-135														

-130														

-125														

Val	Asp	Glu	Ala	Ala	Ala	Ala	Ala	Gly	Asp	Ala	Tyr	Gly	Gly	
-120														

-115														

-110														

Ser	Val	Phe	Asp	Thr	Glu	Thr	Leu	Glu	Leu	Thr	Val	Leu	Val	Thr	Asp
-105															

-100															

-95															

Ala	Ala	Ser	Val	Glu	Ala	Val	Glu	Ala	Thr	Gly	Asp	Leu	
-85													

-80														

-75														

Val	Ser	Tyr	Gly	Ile	Glu	Gly	Leu	Asp	Glu	Ile	Ile	Gln	Asp	Leu	Asn
-70															

-65															

-60															

Ala	Ala	Asp	Ala	Val	Pro	Gly	Val	Val	Gly	Trp	Tyr	Pro	Asp	Val	Ala
-55															

-50															

-45															

Gly	Asp	Thr	Val	Val	Glu	Val	Glu	Gly	Ser	Gly	Ala	Asp	Val	
-40														

-35															

-30															

Ser	Gly	Leu	Leu	Ala	Asp	Ala	Gly	Val	Asp	Ala	Ser
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 30495.204-WO-ST25.txt

Thr Ser Ser Ala Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu  
                   -5                  -1          1                  5

Ala Tyr Thr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn  
                   10                     15                  20

Ala Ala Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Arg Val  
                   25                     30                  35

Gly Thr Gln Val Ser Ile Gly Asn Gln Gly Val Phe Glu Gln Ser  
                   40                     45                  50                  55

Ile Phe Pro Gly Asn Asp Ala Ala Phe Val Arg Gly Thr Ser Asn Phe  
                   60                     65                  70

Thr Leu Thr Asn Leu Val Ser Arg Tyr Asn Thr Gly Gly Tyr Ala Thr  
                   75                     80                  85

Val Ala Gly His Asn Gln Ala Pro Ile Gly Ser Ser Val Cys Arg Ser  
                   90                     95                  100

Gly Ser Thr Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Gly Gln  
                   105                     110                  115

Ser Val Ser Tyr Pro Glu Gly Thr Val Thr Asn Met Thr Arg Thr Thr  
                   120                     125                  130                  135

Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Asn  
                   140                     145                  150

Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Arg Thr Gly  
                   155                     160                  165

Gly Thr Thr Phe Tyr Gln Glu Val Thr Pro Met Val Asn Ser Trp Gly  
                   170                     175                  180

Val Arg Leu Arg Thr  
                   185

<210> 38  
 <211> 43  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Primer 1346

<400> 38  
 gttcatcgat cgcacccggac cgctccccca gtc

<210> 39  
 <211> 38

## 10495.204-WO.ST25.txt

<212> DNA  
<213> Artificial sequence

<220>  
<221> Primer 1602

<400> 39  
gcggatccca ttaggtccgg agacggacgc cccaggag 38

<210> 40  
<211> 1062  
<212> DNA  
<213> Nocardiopsis prasina DSM 15649

<400> 40  
gccaccggac cactcccca gtcacccacc ccggaggccg acggcgctc catgcaggag 60  
gcgcgtccaggc gcgcaccccg cttgagggccg atgaactgtt ggcgcggcc 120  
gacaccgcct tcgaggtcga cgaggcccg gccgaggccg ccggtgacgc ctacggcgcc 180  
tccgtttcg acaccgagac cctggaaactg accgtccctgg tcaccyactc cgccgcggtc 240  
gaggcggtgg aggccaccgg cgccgggacc gaactgggtt cctacggcat cacgggcctc 300  
gacgagatcg tcgaggagct caacgcccgc gacgcgttc cccgggtggc cggctgggtac 360  
ccggacgtcg cgggtgacac cgtcgtgtcg gaggtccctgg agggttcccg cgccgacgtg 420  
ggcggcctgc tcgcccacgc cygcgtggac gcctcggccgg tcgaggtgac caccaccgag 480  
cagccccgagc tgtacyccga catcatcgcc ygtctggcct acaccatggg cggccgtgt 540  
tcggtcggct tcgcccacgc caacgcccgc ggtcagcccg gttcgtcac cgccggtcac 600  
tgtggcccgcg tgcccaccga ggtgaccatc ggcaacggcc ggggcgttcc cgagcagtcc 660  
atcttcccg gcaacgacgc cgccattcgac cgccggacgt ccaacttcac gctgaccaac 720  
ctggtcagcc gctacaacac cggcggtcac gccaccgtcg ccggtcacaa ccaggcgccc 780  
atcggcttcc cggtctggcg ctccgggtcc accaccgggtt ggcactgcgg caccatccag 840  
gccccggcc agtcggtgag ctaccccgag ggcaccgtca ccaacatgac gcgaccacc 900  
gtgtgcgcgg agcccgccga ctccggccggc tcctacatct ccggcaacca ggcccaggc 960  
gtcacctccg ggggtcccg caactgcccgc accggcggtt ccacccatcta ccaggaggc 1020  
accccccattgg tgaactccctg gggcggtccgt ctccggaccc 1062

<210> 41  
<211> 353  
<212> PRT  
<213> Nocardiopsis prasina DSM 15649

<220>  
<221> PROPEP  
<222> (1)..(165)

<220>  
<221> mat\_peptide  
<222> (166)..(353)

<400> 41

## 10495.204-WO-ST25.txt

Ala Thr Gly Pro Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala  
 -165 -160 -155

Val Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Gly Leu Thr Pro  
 -150 -145 -140

Leu Glu Ala Asp Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu  
 -135 -130 -125

Val Asp Glu Ala Ala Ala Glu Ala Ala Gly Asp Ala Tyr Gly Gly  
 -120 -115 -110

Ser Val Phe Asp Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp  
 -105 -100 -95 -90

Ser Ala Ala Val Glu Ala Val Gln Ala Thr Gly Ala Gly Thr Glu Leu  
 -85 -80 -75

Val Ser Tyr Gly Ile Thr Gly Leu Asp Glu Ile Val Glu Glu Leu Asn  
 -70 -65 -60

Ala Ala Asp Ala Val Pro Gly Val Val Gly Trp Tyr Pro Asp Val Ala  
 -55 -50 -45

Gly Asp Thr Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val  
 -40 -35 -30

Gly Gly Leu Leu Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val  
 -25 -20 -15 -10

Thr Thr Thr Glu Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu  
 -5 -1 1 5

Ala Tyr Thr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn  
 10 15 20

Ala Ala Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Arg Val  
 25 30 35

Gly Thr Gln Val Thr Ile Gly Asn Gly Arg Gly Val Phe Glu Gln Ser  
 40 45 50 55

Ile Phe Pro Gly Asn Asp Ala Ala Phe Val Arg Gly Thr Ser Asn Phe  
 60 65 70

Thr Leu Thr Asn Leu Val Ser Arg Tyr Asn Thr Gly Gly Tyr Ala Thr  
 75 80 85

Val Ala Gly His Asn Gln Ala Pro Ile Gly Ser Ser Val Cys Arg Ser  
 90 95 100

10495.204-WO.ST25.txt

Gly Ser Thr Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Gly Gln  
 105                   110                   115

Ser Val Ser Tyr Pro Glu Gly Thr Val Thr Asn Met Thr Arg Thr Thr  
 120                   125                   130                   135

Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Asn  
 140                   145                   150

Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Arg Thr Gly  
 155                   160                   165

Gly Thr Thr Phe Tyr Gln Glu Val Thr Pro Met Val Asn Ser Trp Gly  
 170                   175                   180

Val Arg Leu Arg Thr  
 185

<210> 42  
<211> 43  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Primer 1603

<400> 42  
gttcatcgat cgcacggct gccacggac cactccccca gtc

43

<210> 43  
<211> 353  
<212> PRT  
<213> Nocardiopsis sp. NRRL 18262

<220>  
<221> PROPEP  
<222> (1)..(165)

<220>  
<221> mat\_peptide  
<222> (166)..(1059)

<400> 43

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala  
 -165                   -160                   -155

Val Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser  
 -150                   -145                   -140

Ala Glu Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu  
 -135                   -130                   -125

Val Asp Glu Ala Ala Ala Glu Ala Ala Gly Asp Ala Tyr Gly Gly  
 -120                   -115                   -110

## 10495.204-WO.5T25.txt

Ser Val Phe Asp Thr Glu Ser Leu Glu Leu Thr Val Leu Val Thr Asp  
 -105 -100 -95 -90

Ala Ala Ala Val Glu Ala Val Glu Ala Gly Ala Gly Thr Glu Leu  
 -85 -80 -75

Val Ser Tyr Gly Ile Asp Gly Leu Asp Glu Ile Val Gln Glu Leu Asn  
 -70 -65 -60

Ala Ala Asp Ala Val Pro Gly Val Val Gly Trp Tyr Pro Asp Val Ala  
 -55 -50 -45

Gly Asp Thr Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val  
 -40 -35 -30

Ser Gly Leu Leu Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val  
 -25 -20 -15 -10

Thr Thr Ser Asp Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu  
 -5 -1 1 5

Ala Tyr Thr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn  
 10 15 20

Ala Ala Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Arg Val  
 25 30 35

Gly Thr Gln Val Thr Ile Gly Asn Gly Arg Gly Val Phe Glu Gln Ser  
 40 45 50 55

Val Phe Pro Gly Asn Asp Ala Ala Phe Val Arg Gly Thr Ser Asn Phe  
 60 65 70

Thr Leu Thr Asn Leu Val Ser Arg Tyr Asn Thr Gly Gly Tyr Ala Thr  
 75 80 85

Val Ala Gly His Asn Gln Ala Pro Ile Gly Ser Ser Val Cys Arg Ser  
 90 95 100

Gly Ser Thr Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Gly Gln  
 105 110 115

Ser Val Ser Tyr Pro Glu Gly Thr Val Thr Asn Met Thr Arg Thr Thr  
 120 125 130 135

Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Thr  
 140 145 150

Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Arg Thr Gly  
 155 160 165

## 10495.204-WO.ST25.txt

Gly Thr Thr Phe Tyr Gln Glu Val Thr Pro Met Val Asn Ser Trp Gly  
 170 175 180

Val Arg Leu Arg Thr  
 185

<210> 44  
<211> 1164  
<212> DNA  
<213> Artificial sequence

<220>  
<223> Synthetic protease encoding gene

<220>  
<221> CDS  
<222> (1)..(1164)  
<223> Full length protease

<220>  
<221> sig\_peptide  
<222> (1)..(81)

<220>  
<221> misc\_feature  
<222> (82)..(1164)  
<223> Propeptide

<220>  
<221> mat\_peptide  
<222> (577)..(1164)

<400> 44  
atg aaa aaa ccg ctg gga aaa att gtc gca agc aca gca ctt ctt 45  
Met Lys Lys Pro Leu Gly Lys Ile Val Ala Ser Thr Ala Leu Leu  
-190 -185 -180

att tca gtg gca ttt agc tca tct att gca tca gca gct aca gga 90  
Ile Ser Val Ala Phe Ser Ser Ser Ile Ala Ser Ala Ala Thr Gly  
-175 -170 -165

gca tta ccg cag tct ccg aca ccg gaa gca gat gca gtc tca atg 135  
Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val Ser Met  
-160 -155 -150

caa gaa gca ctg caa aga gat ctt gat ctt aca tca gca gaa gca 180  
Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu Ala  
-145 -140 -135

gaa gaa ctt ctt gct gca caa gat aca gca ttt gaa gtq gat gaa 225  
Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
-130 -125 -120

gca gcg gca gaa gca gca gga gat gca tat ggc ggc tca gtt ttt 270  
Ala Ala Ala Glu Ala Ala Gly Asp Ala Tyr Gly Ser Val Phe  
-115 -110 -105

gat aca gaa tca ctt gaa ctt aca gtt ctt gtt aca gat gca gca gca 318  
Asp Thr Glu Ser Leu Glu Leu Thr Val Leu Val Thr Asp Ala Ala Ala  
-100 -95 -90

gtt gaa gca gtt gaa gca aca gga gca gca gta ctt gtt tca tat 366  
Val Glu Ala Val Glu Ala Thr Gly Ala Gly Thr Val Leu Val Ser Tyr  
-85 -80 -75

## 10495.204-WO.ST25.txt

gga att gat ggc ctt gat gaa att gtt caa gaa ctg aat gca gct gat Gly Ile Asp Gly Leu Asp Glu Ile Val Gln Glu Leu Asn Ala Ala Asp -70 -65 -60 -55	414
gct gtt ccg ggc gtt gtt ggc tgg tat ccg gat gtt gct gga gat aca Ala Val Pro Gly Val Val Gly Trp Tyr Pro Asp Val Ala Gly Asp Thr -50 -45 -40	462
gtt gtc ctt gaa gtt ctt gaa gga tca ggc gca gat gtt tca ggc ctg Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Ser Gly Leu -35 -30 -25	510
ctg gca gac gca gga gtc gat gca tca gca gtt gaa gtt aca aca tca Leu Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val Thr Thr Ser -20 -15 -10	558
gat caa ccg gaa ctt tat gca gat att att ggc ggc ctg gca tat tat Asp Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu Ala Tyr Tyr -5 -1 1 5 10	606
atg ggc ggc aga tgc agc gtt ggc ttt gca gca aca aat gca tca ggc Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn Ala Ser Gly 15 20 25	654
caa ccg ggc ttt gtt aca gca ggc cat tgc ggc aca gtt ggc aca cca Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Thr Val Gly Thr Pro 30 35 40	702
gtt tca att ggc aat ggc aaa ggc gtt ttt gaa cga agc att ttt ccg Val Ser Ile Gly Asn Gly Lys Gly Val Phe Glu Arg Ser Ile Phe Pro 45 50 55	750
ggc aat gat tca gca ttt gtt aga ggc aca tca aat ttt aca ctt aca Gly Asn Asp Ser Ala Phe Val Arg Gly Thr Ser Asn Phe Thr Leu Thr 60 65 70	798
aat ctg gtt tca aga tat aat tca ggc ggc tat gca aca gtt gca ggc Asn Leu Val Ser Arg Tyr Asn Ser Gly Gly Tyr Ala Thr Val Ala Gly 75 80 85 90	846
cat aat caa gca ccg att ggc tca gca gtt tgc aga tca ggc tca aca His Asn Gln Ala Pro Ile Gly Ser Ala Val Cys Arg Ser Gly Ser Thr 95 100 105	894
aca ggc tgg cat tgc ggc aca att caa gca aga aat caa aca gtt agg Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Asn Gln Thr Val Arg 110 115 120	942
tat ccg caa ggc aca gtt tat agt ctg aca aga aca aca gtt tgt gca Tyr Pro Gln Gly Thr Val Tyr Ser Leu Thr Arg Thr Val Cys Ala 125 130 135	990
gaa ccg ggc gat tca ggc ggc tca tat att agc ggc act caa gca caa Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln 140 145 150	1038
ggc gtt aca tca ggc ggc tca ggc aat tgc agt gct ggc ggc aca aca Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Ser Ala Gly Gly Thr Thr 155 160 165 170	1086
tat tac caa gaa gtt aat ccg atg ctt agt tca tgg ggc ctt aca ctt Tyr Tyr Gln Glu Val Asn Pro Met Leu Ser Ser Trp Gly Leu Thr Leu 175 180 185	1134
aga aca caa tcg cat gtt caa tcc gct cca Arg Thr Gln Ser His Val Gln Ser Ala Pro 190 195	1164

## 10495.204-WO.ST25.txt

<210> 45  
<211> 388  
<212> PRT  
<213> Artificial sequence

<220>  
<223> Synthetic Construct

<400> 45

Met Lys Lys Pro Leu Gly Lys Ile Val Ala Ser Thr Ala Leu Leu  
-190 -185 -180

Ile Ser Val Ala Phe Ser Ser Ser Ile Ala Ser Ala Ala Thr Gly  
-175 -170 -165

Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val Ser Met  
-160 -155 -150

Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu Ala  
-145 -140 -135

Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
-130 -125 -120

Ala Ala Ala Glu Ala Ala Gly Asp Ala Tyr Gly Gly Ser Val Phe  
-115 -110 -105

Asp Thr Glu Ser Leu Glu Leu Thr Val Leu Val Thr Asp Ala Ala Ala  
-100 -95 -90

Val Glu Ala Val Glu Ala Thr Gly Ala Gly Thr Val Leu Val Ser Tyr  
-85 -80 -75

Gly Ile Asp Gly Leu Asp Glu Ile Val Gln Glu Leu Asn Ala Ala Asp  
-70 -65 -60 -55

Ala Val Pro Gly Val Val Gly Trp Tyr Pro Asp Val Ala Gly Asp Thr  
-50 -45 -40

Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Ser Gly Leu  
-35 -30 -25

Leu Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val Thr Thr Ser  
-20 -15 -10

Asp Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu Ala Tyr Tyr  
-5 -1 1 5 10

Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn Ala Ser Gly  
15 20 25

10495\_204-WO\_ST25.txt  
 Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Thr Val Gly Thr Pro  
 30 35 40

Val Ser Ile Gly Asn Gly Lys Gly Val Phe Glu Arg Ser Ile Phe Pro  
 45 50 55

Gly Asn Asp Ser Ala Phe Val Arg Gly Thr Ser Asn Phe Thr Leu Thr  
 60 65 70

Asn Leu Val Ser Arg Tyr Asn Ser Gly Gly Tyr Ala Thr Val Ala Gly  
 75 80 85 90

His Asn Gln Ala Pro Ile Gly Ser Ala Val Cys Arg Ser Gly Ser Thr  
 95 100 105

Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Asn Gln Thr Val Arg  
 110 115 120

Tyr Pro Gln Gly Thr Val Tyr Ser Leu Thr Arg Thr Thr Val Cys Ala  
 125 130 135

Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln  
 140 145 150

Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Ser Ala Gly Gly Thr Thr  
 155 160 165 170

Tyr Tyr Gln Glu Val Asn Pro Met Leu Ser Ser Trp Gly Leu Thr Leu  
 175 180 185

Arg Thr Gln Ser His Val Gln Ser Ala Pro  
 190 195

<210> 46  
 <211> 165  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> shuffled propeptide O-2.19

<220>  
 <221> PROPEP  
 <222> (1)..(165)

<400> 46

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val  
 1 5 10 15

Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu  
 20 25 30

Ala Glu Gln Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
 Page 30

30495.204-WO-ST25.txt  
 35                          40                          45

Ala Ala Ala Ala Ala Ala Gly Asp Ala Tyr Gly Gly Ser Val Phe Asp  
 50                          55                          60

Thr Glu Ser Leu Thr Leu Thr Val Leu Val Thr Asp Ala Ser Ala Val  
 65                          70                          75                          80

Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly  
 85                          90                          95

Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala  
 100                        105                        110

Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val  
 115                        120                        125

Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu  
 130                        135                        140

Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val Thr Thr Ser Asp  
 145                        150                        155                        160

Gln Pro Glu Leu Tyr  
 165

<210> 47  
<211> 166  
<212> PRT  
<213> Artificial sequence

<220>  
<223> shuffled propeptide G-2.73

<220>  
<221> PROPEP  
<222> (1)..(166)

<400> 47

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val  
 1                        5                            10                        15

Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Ser Ser Ala Glu  
 20                        25                        30

Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
 35                        40                        45

Ala Ala Ala Gly Ala Ala Gly Asp Ala Tyr Gly Gly Ser Val Phe Asp  
 50                        55                        60

Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp Ala Ser Ala Val  
 65                        70                        75                        80

10495.204-WO,ST25.txt

Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly  
 85 90 95

Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala  
 100 105 110

Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val  
 115 120 125

Val Val Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu  
 130 135 140

Ala Asp Ala Gly Val Asp Thr Ala Asp Val Lys Val Glu Ser Thr Thr  
 145 150 155 160

Glu Gln Pro Glu Leu Tyr  
 165

<210> 48

<211> 166

<212> PRT

<213> Artificial sequence

<220>

<223> shuffled propeptide G-1,43

<220>

<221> PROPEP

<222> (1)..(166)

<400> 48

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val  
 1 5 10 15

Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Gly Leu Ser Ser Ser Gln  
 20 25 30

Ala Glu Glu Leu Leu Asp Ala Gln Ala Glu Ser Phe Glu Ile Asp Glu  
 35 40 45

Ala Ala Ala Ala Ala Ala Gly Asp Ala Tyr Gly Gly Ser Ile Phe Asp  
 50 55 60

Thr Asp Ser Leu Thr Leu Thr Val Leu Val Thr Asp Ala Ser Ala Val  
 65 70 75 80

Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly  
 85 90 95

Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala  
 100 105 110

10495, 204-W0, ST25, txt

Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val  
115 120 125

Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu  
130 135 140

Ala Asp Ala Gly Val Asp Thr Ala Asp Val Lys Val Glu Ser Thr Thr  
145 150 155 160

Glu Gln Pro Glu Leu Tyr  
16S

<210> 49  
<211> 166  
<212> PRT  
<213> Artificial sequence

<220>  
<223> shuffled propeptide G-2.6

4400> 49

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val  
1 5 10 15

Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu  
20 25 30

Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
35 40 45

Ala Ala Ala Ala Ala Gly Asp Ala Tyr Gly Gly Ser Ile Phe Asp  
50 55 60

Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp Ser Ser Ser Val  
65 70 75 80

Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly  
85 90 95

Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala  
100 105 110

Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val  
115 120 125

Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu  
130 135 140

Ala Gly Ala Gly Val Asp Thr Ala Asp Val Lys Val Glu Ser Thr Thr  
145 150 155 160

Glu Gln Pro Glu Leu Tyr

10495.204-WO.ST25.txt

165

<210> 50  
<211> 165  
<212> PRT  
<213> Artificial sequence

<220>  
<223> Shuffled propeptide G-2.5

<220>  
<221> PROPEP  
<222> (1)..(165)

&lt;400&gt; 50

Ala	Thr	Gly	Ala	Leu	Pro	Gln	Ser	Pro	Thr	Pro	Glu	Ala	Asp	Ala	Val
1				5				10				15			

Ser	Met	Gln	Glu	Ala	Leu	Gln	Arg	Asp	Leu	Gly	Leu	Thr	Pro	Leu	Glu
						25						30			

Ala	Glu	Glu	Leu	Leu	Ala	Ala	Gln	Asp	Thr	Ala	Phe	Glu	Val	Asp	Glu
					35		40					45			

Ala	Ala	Ala	Glu	Ala	Ala	Gly	Asp	Ala	Tyr	Gly	Gly	Ser	Val	Phe	Asp
					50	55					60				

Thr	Glu	Thr	Leu	Glu	Leu	Thr	Val	Leu	Val	Thr	Asp	Ala	Ser	Ala	Val
					65		70			75		80			

Glu	Ala	Val	Glu	Ala	Ala	Gly	Ala	Glu	Ala	Lys	Val	Val	Ser	His	Gly
					85		90				95				

Met	Glu	Gly	Leu	Glu	Glu	Ile	Val	Ala	Asp	Leu	Asn	Ala	Ala	Asp	Ala
					100		105				110				

Gln	Pro	Gly	Val	Val	Gly	Trp	Tyr	Pro	Asp	Ile	Nis	Ser	Asp	Thr	Val
					115		120				125				

Val	Leu	Glu	Val	Leu	Glu	Gly	Ser	Gly	Ala	Asp	Val	Asp	Ser	Leu	Leu
					130		135				140				

Ala	Asp	Ala	Gly	Val	Asp	Ala	Ser	Ala	Val	Glu	Val	Thr	Pro	Ala	Ala
					145		150			155			160		

Arg Pro Glu Leu Tyr  
165

<210> 51  
<211> 166  
<212> PRT  
<213> Artificial sequence

&lt;220&gt;

10495.204-WO.ST25.txt

&lt;223&gt; shuffled propeptide G-2.3

&lt;220&gt;

&lt;221&gt; PROPEP

&lt;222&gt; (1)..(166)

&lt;400&gt; 51

Ala	Thr	Gly	Ala	Leu	Pro	Gln	Ser	Pro	Thr	Pro	Asp	Gly	Ala	Glu	Ala
1				5					10				15		

Thr	Thr	Met	Val	Glu	Ala	Leu	Gln	Arg	Asp	Leu	Gly	Leu	Thr	Pro	Ala
				20				25				30			

Glu	Ala	Glu	Glu	Leu	Leu	Ala	Ala	Gln	Asp	Thr	Ala	Phe	Glu	Val	Asp
		35				40						45			

Glu	Ala	Ala	Ala	Ala	Ala	Gly	Asp	Ala	Tyr	Gly	Gly	Ser	Ile	Phe
		50				55				60				

Asp	Thr	Asp	Ser	Leu	Thr	Leu	Thr	Val	Leu	Val	Thr	Asp	Ala	Ala	Ala
				65			70			75		80			

Val	Glu	Ala	Val	Glu	Ala	Ala	Gly	Ala	Glu	Ala	Lys	Val	Val	Ser	His
				85				90			95				

Gly	Met	Glu	Gly	Leu	Glu	Glu	Ile	Val	Ala	Asp	Leu	Asn	Ala	Ala	Asp
				100			105					110			

Ala	Val	Pro	Gly	Val	Val	Gly	Trp	Tyr	Pro	Asp	Val	Ala	Gly	Asp	Thr
				115			120					125			

Val	Val	Leu	Glu	Val	Leu	Glu	Gly	Ser	Gly	Ala	Asp	Val	Tyr	Ser	Leu
		130				135			140						

Leu	Ala	Asp	Ala	Gly	Val	Asp	Ala	Ser	Ala	Val	Glu	Val	Thr	Pro	Ala
				145				150			155			160	

Ala	Gln	Pro	Glu	Leu	Tyr										
				165											

&lt;210&gt; 52

&lt;211&gt; 166

&lt;212&gt; PRT

&lt;213&gt; Artificial sequence

<220>															
<223>	shuffled propeptide	G-1.4													

&lt;220&gt;

&lt;221&gt; PROPEP

&lt;222&gt; (1)..(166)

&lt;400&gt; 52

10495\_204-WO\_ST25.txt

Ala	Thr	Gly	Ala	Leu	Pro	Gln	Ser	Pro	Thr	Pro	Glu	Ala	Asp	Ala	Val
1															15
Ser	Met	Gln	Glu	Ala	Leu	Gln	Arg	Asp	Leu	Gly	Ieu	Ser	Ser	Ser	Gln
															30
Ala	Glu	Glu	Leu	Leu	Asp	Ala	Gln	Ala	Glu	Ser	Phe	Glu	Ile	Asp	Glu
															45
Ala	Ala	Ala	Ala	Ala	Ala	Asp	Ser	Tyr	Gly	Gly	Ser	Ile	Phe	Asp	
															60
Thr	Asp	Ser	Leu	Thr	Leu	Thr	Val	Leu	Val	Thr	Asp	Ala	Ser	Ala	Val
															80
Glu	Ala	Val	Glu	Ala	Ala	Gly	Ala	Glu	Ala	Lys	Val	Val	Ser	His	Gly
															95
Met	Glu	Gly	Leu	Glu	Glu	Ile	Val	Ala	Asp	Leu	Asn	Ala	Ala	Asp	Ala
															110
Gln	Pro	Gly	Val	Val	Gly	Trp	Tyr	Pro	Asp	Ile	His	Ser	Asp	Thr	Val
															125
Val	Leu	Glu	Val	Leu	Glu	Gly	Ser	Gly	Ala	Asp	Val	Asp	Ser	Leu	Leu
															140
Ala	Asp	Ala	Gly	Val	Asp	Thr	Ala	Asp	Val	Lys	Val	Glu	Ser	Thr	Thr
															160
Glu	Gln	Pro	Glu	Leu	Tyr										

<210> 53  
<211> 166  
<212> PRT  
<213> Artificial sequence

<220>  
<223> shuffled propeptide G-1,2

<220>  
<221> PROPEP  
<222> (1)..(166)  
<400> 53

Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val  
1 5 10 15

Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu  
20 25 30

Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu  
Page 36

35 30495\_204-WO.ST25.txt  
40 45

Ala Ala Ala Ala Ala Ala Gly Asp Ala Tyr Gly Gly Ser Ile Phe Asp  
50 55 60

Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp Ser Ser Ser Val  
65 70 75 80

Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly  
85 90 95

Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala  
100 105 110

Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val  
115 120 125

Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu  
130 135 140

Ala Gly Ala Gly Val Asp Thr Ala Asp Val Lys Val Glu Ser Thr Thr  
145 150 155 160

Glu Gln Pro Glu Leu Tyr  
165

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/DK2004/000436

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12N9/58

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (names of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, PAD, EMBL, Sequence Search, EMBASE, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passage	Relevant to claim no.
X	DATABASE GSP DERWENT; 19 May 1999 (1999-05-19), NOVO-NORDISK AS.: "Nocardiopsis sp. proteolytic protein." XP002298820 Database accession no. AAH92997 the whole document	1-3,8,9, 16,21
X	-& DATABASE GSN DERWENT; 19 May 1999 (1999-05-19), NOVO-NORDISK AS.: "Nocardiopsis sp. proteolytic protein cDNA." XP002298821 Database accession no. AAX22316 the whole document	10-13
X	& DK 1 396 A (NOVONORDISK AS) 9 January 1996 (1996-01-09)	1-3,8-13
	----- -/-/-	

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the International filing date
- \*L\* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation, or other special reason [as specified]
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the International filing date but later than the priority date claimed

\*T\* later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*&\* document member of the same patent family

Date of the actual completion of the International search

5 October 2004

Date of mailing of the International search report

21/10/2004

## Name and mailing address of the ISA

European Patent Office, P.O. 5610 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel: (+31-70) 340-2040, Tx. 21 851 epo nl,  
Fax: (+31-70) 340-3716

## Authorized officer

Steffen, P

## INTERNATIONAL SEARCH REPORT

International Application No.

T/DK2004/000436

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation or document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/58276 A (HOFFMANN LA ROCHE ; OESTERGAARD PETER RAHBEK (DK); SJØEHOLM CARSTEN (D) 16 August 2001 (2001-08-16) SEQ ID NO's 1 & 2examples 1,2,6	1-3,8,9, 16-18,20
X	MITSUIKI S ET AL: "Purification and some properties of a keratinolytic enzyme from an alkaliphilic Nocardiopsis sp. TOA-1" BIOSCIENCE BIOTECHNOLOGY BIOCHEMISTRY, JAPAN SOC. FOR BIOSCIENCE, BIOTECHNOLOGY AND AGROCHEM. TOKYO, JP, vol. 66, no. 1, January 2002 (2002-01), pages 164-167, XP002296335 ISSN: 0916-8451 page 164 ~ page 166; figure 2	1-3,8,9, 16
P,X	-& MITSUIKI SHINJI ET AL: "Molecular characterization of a keratinolytic enzyme from an alkaliphilic Nocardiopsis sp. TOA-1" ENZYME AND MICROBIAL TECHNOLOGY, vol. 34, no. 5, 2 April 2004 (2004-04-02), pages 482-489, XP002298817 ISSN: 0141-0229	1-3,8,9, 16
X	TSUJIBO H ET AL: "AMINO ACID COMPOSITIONS AND PARTIAL SEQUENCES OF TWO TYPES OF ALKALINE SERINE PROTEASES FROM NOCARDIOPSIS DASSONVILLEI SUBSP. PRASINA OPC.210" AGRICULTURAL AND BIOLOGICAL CHEMISTRY, JAPAN SOC. FOR BIOSCIENCE, BIOTECHNOLOGY AND AGROCHEM. TOKYO, JP, vol. 54, no. 8, 1990, pages 2177-2179, XP001189697 ISSN: 0002-1369 page 2177; figures 1,2	1-3,8,9, 16
X	DIXIT V S ET AL: "Comparative characterization of two serine endopeptidases from Nocardiopsis sp. NCIM 5124" BBA - GENERAL SUBJECTS, ELSEVIER SCIENCE PUBLISHERS, NL, vol. 1523, no. 2-3, 18 October 2000 (2000-10-18), pages 261-268, XP004275897 ISSN: 0304-4165 page 262 ~ page 263; figures 2,5	1-3,8,16

-/-

## INTERNATIONAL SEARCH REPORT

International Application No

P/DK2004/000436

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TAKAHASHI MASAKAZU ET AL: "Improved autoprocessing efficiency of mutant subtilisins E with altered specificity by engineering of the pro-region" JOURNAL OF BIOCHEMISTRY (TOKYO), vol. 130, no. 1, July 2001 (2001-07), pages 99-106, XP009037390 ISSN: 0021-924X the whole document	
A	YAMASHIRO K ET AL: "MOLECULAR CLONING OF A NOVEL TRYPSIN-LIKE SERINE PROTEASE(NEUROSIN)PREFERENTIALLY EXPRESSED IN BRAIN" BIOCHIMICA ET BIOPHYSICA ACTA, AMSTERDAM, NL, vol. 1350, 1997, pages 11-14, XP001056162 ISSN: 0006-3002 the whole document	
A	TANG BING ET AL: "General function of N-terminal propeptide on assisting protein folding and inhibiting catalytic activity based on observations with a chimeric thermolysin-like protease." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 301, no. 4, 21 February 2003 (2003-02-21), pages 1093-1098, XP002298818 ISSN: 0006-291X the whole document	
A	LAO GUIFANG ET AL: "Cloning, sequencing, and expression of a Thermomonospora fusca protease gene in Streptomyces lividans" APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 62, no. 11, 1996, pages 4256-4259, XP002285856 ISSN: 0099-2240 the whole document	
P,X	JP 2003 284571 A (TOTO LTD; DAIWA KASEI KK) 7 October 2003 (2003-10-07) SEQ ID NO's 1 & 2 the whole document	1-3,8,9, 16,21
E	WO 2004/072221 A (DE MARIA LEONARDO ; NOVOZYME AS (DK); OESTERGAARD PETER RAHBEK (DK)) 26 August 2004 (2004-08-26) SEQ ID NO's 5-12claims 11-24	1-3,8-21
E	WO 2004/072279 A (LAN TANG ; LIU YE (CN); WU WENPING (CN); DE MARIA LEONARDO (DK); PONTO) 26 August 2004 (2004-08-26) SEQ ID NO's: 25 & 26claims 14-26	1-3,8-21

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

JP/DK2004/000436

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
DK 1396	A	09-01-1996	NONE			
WO 0158276	A	16-08-2001	AU AU BR BR CA CA CN CN WO WO EP EP JP JP PL PL US US US	3544601 A 4236601 A 0108164 A 0108165 A 2395266 A1 2395343 A1 1398161 T 1398162 T 0158275 A2 0158276 A2 1257175 A2 1257176 A2 2003521907 T 2003521908 T 357638 A1 357668 A1 2004161448 A1 2001026797 A1 2003021774 A1	20-08-2001 20-08-2001 21-01-2003 25-02-2003 16-08-2001 16-08-2001 19-02-2003 19-02-2003 16-08-2001 16-08-2001 20-11-2002 20-11-2002 22-07-2003 22-07-2003 26-07-2004 26-07-2004 19-08-2004 04-10-2001 30-01-2003	
JP 2003284571	A	07-10-2003	NONE			
WO 2004072221	A	26-08-2004	WO WO	2004072221 A2 2004072279 A2	26-08-2004 26-08-2004	
WO 2004072279	A	26-08-2004	WO WO	2004072221 A2 2004072279 A2	26-08-2004 26-08-2004	